

PATENT COOPERATION TREATY

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

From the INTERNATIONAL BUREAU

To:

Assistant Commissioner for Patents
 United States Patent and Trademark
 Office
 Box PCT
 Washington, D.C. 20231
 ÉTATS-UNIS D'AMÉRIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 30 November 1999 (30.11.99)	
International application No. PCT/DE99/01323	Applicant's or agent's file reference GR 98P1687P
International filing date (day/month/year) 03 May 1999 (03.05.99)	Priority date (day/month/year) 11 May 1998 (11.05.98)
Applicant BUB, Udo et al	

1. The designated Office is hereby notified of its election made:

☒ in the demand filed with the International Preliminary Examining Authority on:
 02 November 1999 (02.11.99)

☐ in a notice effecting later election filed with the International Bureau on:

2. The election ☒ was
☐ was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No.: (41-22) 740.14.35	Authorized officer R. Forax Telephone No.: (41-22) 338.83.38
---	--

BEST AVAILABLE COPY

PCT ANTRAG

Original (für EINREICHUNG) - gedruckt am 09.04.1999 01:23:04 PM

0	Vom Anmeldeamt auszufüllen	
0-1	Internationales Aktenzeichen.	
0-2	Internationales Anmeldedatum	
0-3	Name des Anmeldeamts und "PCT International Application"	
0-4 0-4-1	Formular - PCT/RO/101 PCT Antrag erstellt durch Benutzung von	PCT-EASY Version 2.83 (aktualisiert 01.03.1999)
0-5	Antragssersuchen Der Unterzeichnete beantragt, daß die vorliegende internationale Anmeldung nach dem Vertrag über die internationale Zusammenarbeit auf dem Gebiet des Patentwesens behandelt wird	
0-6	(Vom Anmelder gewähltes) Anmeldeamt	Deutsches Patent- und Markenamt (RO/DE)
0-7	Aktenzeichen des Anmelders oder Anwalts	GR 98P1687P
I	Bezeichnung der Erfindung	ANORDNUNG UND VERFAHREN ZUR ERKENNUNG EINES VORGEGEBENEN WORTSCHATZES IN GESPROCHENER SPRACHE DURCH EINEN RECHNER
II	Anmelder	nur Anmelder
II-1	Diese Person ist	Alle Bestimmungstaaten mit Ausnahme von
II-2	Anmelder für	US
II-4	Name	SIEMENS AKTIENGESSELLSCHAFT
II-5	Anschrift:	Wittelsbacherplatz 2 D-80333 München Germany
II-6	Staatsangehörigkeit (Staat)	DE
II-7	Sitz/Wohnsitz (Staat)	DE
II-8	Telefonnr.	(089) 636-82819
II-9	Telefaxnr.	(089) 636-81857
III-1	Anmelder und/oder Erfinder	Anmelder und Erfinder
III-1-1	Diese Person ist	Nur US
III-1-2	Anmelder für	BUB, Udo
III-1-4	Name (FAMILIENNAME, Vorname)	Klarweinstr. 18
III-1-5	Anschrift:	D-81247 München Germany
III-1-6	Staatsangehörigkeit (Staat)	DE
III-1-7	Sitz/Wohnsitz (Staat)	DE

Fr. 7p

PCT ANTRAG

GR 98P1687P

Original (für EINREICHUNG) - gedruckt am 09.04.1999 01:23:04 PM

III-2	Anmelder und/oder Erfinder	
III-2-1	Diese Person ist	Anmelder und Erfinder
III-2-2	Anmelder für	Nur US
III-2-4	Name (FAMILIENNAME, Vorname)	HÖGE, Harald
III-2-5	Anschrift:	Obertaxetweg 6B D-82131 Gauting Germany
III-2-6	Staatsangehörigkeit (Staat)	DE
III-2-7	Sitz/Wohnsitz (Staat)	DE
IV-1	Anwalt oder gemeinsamer Vertreter; oder besondere Zustellanschrift Die unten bezeichnete Person ist/wird hiermit bestellt, um den (die) Anmelder vor den internationalen Behörden zu vertreten, und zwar als:	gemeinsamer Vertreter
IV-1-1	Name	SIEMENS AKTIENGESELLSCHAFT
IV-1-2	Anschrift:	Postfach 22 16 34 D-80506 München Germany
IV-1-3	Telefonnr.	(089) 636-82819
IV-1-4	Telefaxnr.	(089) 636-81857
V	Bestimmung von Staaten	
V-1	Regionales Patent (andere Schutzrechtsarten oder Verfahren sind ggf. in Klammern nach der (den) betreffenden Bestimmung(en) angegeben)	EP: AT BE CH&LI CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE und jeder weitere Staat, der Mitgliedsstaat des Europäischen Patentübereinkommens und Vertragsstaat des PCT ist
V-2	Nationales Patent (andere Schutzrechtsarten oder Verfahren sind ggf. in Klammern nach der (den) betreffenden Bestimmung(en) angegeben)	US
V-5	Erklärung bzgl. vorsorglicher Bestimmungen Zusätzlich zu den unter Punkt V-1, V-2 and V-3 vorgenommenen Bestimmungen nimmt der Anmelder nach Regel 4.9 Absatz b auch alle anderen nach dem PCT zulässigen Bestimmungen vor mit Ausnahme der nachstehend unter Punkt V-6 angegebenen Staaten. Der Anmelder erklärt, daß diese zusätzlichen Bestimmungen unter dem Vorbehalt einer Bestätigung stehen und jede zusätzliche Bestimmung, die vor Ablauf von 15 Monaten ab dem Prioritätsdatum nicht bestätigt wurde, nach Ablauf dieser Frist als vom Anmelder zurückgenommen gilt.	
V-6	Staaten, die von der Erklärung über vorsorgliche Bestimmungen ausgenommen werden	KEINE

PCT ANTRAG

Original (für EINREICHUNG) - gedruckt am 09.04.1999 01:23:04 PM

VI-1	Priorität einer früheren nationalen Anmeldung beansprucht		
VI-1-1	Anmeldedatum	11 Mai 1998 (11.05.1998)	
VI-1-2	Aktenzeichen	198 21 057.4	
VI-1-3	Staat	DE	
VI-2	Ersuchen um Erstellung eines Prioritätsbeleges Das Anmeldeamt wird ersucht, eine beglaubigte Abschrift der in der (den) nachstehend genannten Zeile(n) bezeichneten früheren Anmeldung(en) zu erstellen und dem internationalen Büro zu übermitteln:	VI-1	
VII-1	Gewählte internationale Recherchenbehörde	Europäisches Patentamt (EPA) (ISA/EP)	
VIII	Kontrollliste	Anzahl der Blätter	Elektronische Datei(en) beigelegt
VIII-1	Antrag	4	-
VIII-2	Beschreibung	14	-
VIII-3	Ansprüche	3	-
VIII-4	Zusammenfassung	1	98_p_1687_p_de.txt
VIII-5	Zeichnung(en)	1	-
VIII-7	INSGESAMT	23	
	Beigelegte Unterlagen	Unterlage(n) in Papierform beigelegt	Elektronische Datei(en) beigelegt
VIII-8	Blatt für die Gebührenberechnung	✓	-
VIII-16	PCT-EASY-Diskette	-	Diskette
VIII-17	Sonstige (einzeln aufgeführt):	Kopie der Ursprungsfassung	-
VIII-18	Nr. der Abb. der Zeichn., die mit der Zusammenf. veröffentlicht werden soll	1	
VIII-19	Sprache der int. Anmeldung	Deutsch	
IX-1	Unterschrift des Anmelders oder Anwalts		
IX-1-1	Name	i.v. Haslunbe SIEMENS AKTIENGESELLSCHAFT	
IX-1-2	Name der unterzeichnenden Person	Hashuber	
IX-1-3	Eigenschaft	Nr. 144/74 Ang-AV	
IX-2	Unterschrift des Anmelders oder Anwalts		
IX-2-1	Name	SIEMENS AKTIENGESELLSCHAFT	
IX-3	Unterschrift des Anmelders oder Anwalts	Udo Bub	
IX-3-1	Name (FAMILIENNAME, Vorname)	BUB, Udo	
IX-4	Unterschrift des Anmelders oder Anwalts	Harald Höge	
IX-4-1	Name (FAMILIENNAME, Vorname)	HÖGE, Harald	

VOM ANMELDEAMT AUSZUFÜLLEN

10-1	Datum des tatsächlichen Eingangs dieser internationalen Anmeldung	
10-2	Zeichnung(en):	
10-2-1	Eingegangen	
10-2-2	Nicht eingegangen	
10-3	Geändertes Eingangsdatum aufgrund nachträglich, jedoch fristgerecht eingeg. Unterlage(n) oder Zeichnung(en) zur Vervollständigung dieser int. Anmeldung	
10-4	Datum des fristgerechten Eingangs der Berichtigung nach PCT Artikel 11(2)	
10-5	Internationale Recherchenbehörde	ISA/EP
10-6	Übermittlung des Recherchenexemplars bis zur Zahlung der Recherchegebühr aufgeschoben	

VOM INTERNATIONALEN BÜRO AUSZUFÜLLEN

11-1	Datum des Eingangs des Aktenexemplars beim Internationalen Büro	
------	---	--

**PCT (ANHANG - BLATT FÜR DIE
GEBÜHRENBERECHNUNG)**

Original (für EINREICHUNG) - gedruckt am 09.04.1999 01:23:04 PM

(Dieses Blatt ist nicht Teil und zählt nicht als Blatt der internationalen Anmeldung)

0	Vom Anmeldeamt auszufüllen		
0-1	Internationales Aktenzeichen.		
0-2	Eingangsstempel des Anmeldeamts		
0-4	Formular - PCT/RO/101 (Anlage) PCT Blatt für die Gebührenberechnung erstellt durch Benutzung von	PCT-EASY Version 2.83 (aktualisiert 01.03.1999)	
0-9	Aktenzeichen des Anmelders oder Anwalts	GR 98P1687P	
2	Anmelder	SIEMENS AKTIENGESELLSCHAFT, et al.	
12	Berechnung der vorgeschriebenen Gebühren	Höhe der Gebühr/Multiplikator	Gesamtbeträge (DEM)
12-1	Übermittlungsgebühr T	⇒	150
12-2	Recherchegebühr S	⇒	2.198,35
12-3	Internationale Gebühr Grundgebühr (erste 30 Blätter) b1	800	
12-4	Remaining sheets	0	
12-5	Zusatzblattgebühr (X)	19	
12-6	Gesamtbetrag der weiteren Gebühren b2	0	
12-7	b1 + b2 = B	800	
12-8	Bestimmungsgebühren Anzahl der in der internationalen Anmeldung vorgenommenen Bestimmungen	2	
12-9	Anzahl der zu zahlenden Bestimmungsgebühren (höchstens 10)	2	
12-10	Bestimmungsgebühr (X)	184	
12-11	Gesamtbetrag der Bestimmungsgebühren D	368	
12-12	PCT-EASY-Gebührenermäßigung R	-246	
12-13	Gesamtbetrag der internationalen Gebühr (B+D+R) I	⇒	922
12-14	Gebühr für Prioritätsbeleg Anzahl der beantragten Prioritätsbelege	1	
12-15	Gebühr per Prioritätsbeleg (X)	35	
12-16	Gesamtbetrag Gebühr für Prioritätsbeleg(e) P	⇒	35
12-17	GESAMTBETRAG DER ZU ZAHLENDEN GEBÜHREN (T+S+I+P)	⇒	3.305,35
12-18	Zahlungsart	Sonstige: Abbuchung durch gesonderte Zahlungsliste	

**PCT (ANHANG - BLATT FÜR DIE
GEBÜHRENBERECHNUNG)**

Original (für EINREICHUNG) - gedruckt am 09.04.1999 01:23:04 PM

12-20	Anweisungen betreffend laufendes Konto Das Anmeldeamt:	Deutsches Patent- und Markenamt (RO/DE)
12-20-2	wird beauftragt, Fehlbeträge oder Überzahlungen des vorstehend angegebenen Gesamtbetrags der Gebühren meinem laufenden Konto zu belasten bzw. gutzuschreiben	✓
12-21	Nummer des laufenden Kontos	409022601
12-22	Datum	09 April 1999 (09.04.1999)
12-23	Name und Unterschrift	SIEMENS AKTIENGESELLSCHAFT <i>i. V. Haslube</i>

PRÜFPROTOKOLL UND BEMERKUNGEN

13-2-1	Prüfergebnisse Antrag	Grün? Die Bezeichnung der Erfindung muß kurz und genau gefaßt sein. Bitte überprüfen.
13-2-2	Prüfergebnisse Staaten	Grün? Es können mehr Bestimmungen vorgenommen werden. Bitte überprüfen.

VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS

PCT

INTERNATIONALER VORLÄUFIGER PRÜFUNGSBERICHT

(Artikel 36 und Regel 70 PCT)

REC'D 11 AUG 2000

WIPO

PCT

Aktenzeichen des Anmelders oder Anwalts GR 98P1687P	WEITERES VORGEHEN siehe Mitteilung über die Übersendung des internationalen vorläufigen Prüfungsbericht (Formblatt PCT/IPEA/416)	
Internationales Aktenzeichen PCT/DE99/01323	Internationales Anmeldedatum (Tag/Monat/Jahr) 03/05/1999	Prioritätsdatum (Tag/Monat/Tag) 11/05/1998
Internationale Patentklassifikation (IPK) oder nationale Klassifikation und IPK G10L5/06		
Anmelder SIEMENS AKTIENGESELLSCHAFT et al.		

1. Dieser internationale vorläufige Prüfungsbericht wurde von der mit der internationale vorläufigen Prüfung beauftragte Behörde erstellt und wird dem Anmelder gemäß Artikel 36 übermittelt.


2. Dieser BERICHT umfaßt insgesamt 4 Blätter einschließlich dieses Deckblatts.

☒ Außerdem liegen dem Bericht ANLAGEN bei; dabei handelt es sich um Blätter mit Beschreibungen, Ansprüchen und/oder Zeichnungen, die geändert wurden und diesem Bericht zugrunde liegen, und/oder Blätter mit vor dieser Behörde vorgenommenen Berichtigungen (siehe Regel 70.16 und Abschnitt 607 der Verwaltungsrichtlinien zum PCT).

Diese Anlagen umfassen insgesamt 6 Blätter.

3. Dieser Bericht enthält Angaben zu folgenden Punkten:

- I ☒ Grundlage des Berichts
- II ☐ Priorität
- III ☐ Keine Erstellung eines Gutachtens über Neuheit, erfinderische Tätigkeit und gewerbliche Anwendbarkeit
- IV ☐ Mangelnde Einheitlichkeit der Erfindung
- V ☒ Begründete Feststellung nach Artikel 35(2) hinsichtlich der Neuheit, der erfinderische Tätigkeit und der gewerbliche Anwendbarkeit; Unterlagen und Erklärungen zur Stützung dieser Feststellung
- VI ☐ Bestimmte angeführte Unterlagen
- VII ☐ Bestimmte Mängel der internationalen Anmeldung
- VIII ☐ Bestimmte Bemerkungen zur internationalen Anmeldung

Datum der Einreichung des Antrags 02/11/1999	Datum der Fertigstellung dieses Berichts 09.08.2000
Name und Postanschrift der mit der internationalen vorläufigen Prüfung beauftragten Behörde:  Europäisches Patentamt D-80298 München Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Bevollmächtigter Bediensteter Greiser, N Tel. Nr. +49 89 2399 7402



I. Grundlage des Berichts

1. Dieser Bericht wurde erstellt auf der Grundlage (*Ersatzblätter, die dem Anmeldeamt auf eine Aufforderung nach Artikel 14 hin vorgelegt wurden, gelten im Rahmen dieses Berichts als "ursprünglich eingereicht" und sind ihm nicht beigefügt, weil sie keine Änderungen enthalten.*):

Beschreibung, Seiten:

1-3,5-14 ursprüngliche Fassung

4,4a eingegangen am 07/07/2000 mit Schreiben vom 03/07/2000

Patentansprüche, Nr.:

1-8 eingegangen am 07/07/2000 mit Schreiben vom 03/07/2000

Zeichnungen, Blätter:

1/1 ursprüngliche Fassung

2. Aufgrund der Änderungen sind folgende Unterlagen fortgefallen:

- ☐ Beschreibung, Seiten:
- ☐ Ansprüche, Nr.:
- ☐ Zeichnungen, Blatt:

3. ☐ Dieser Bericht ist ohne Berücksichtigung (von einigen) der Änderungen erstellt worden, da diese aus den angegebenen Gründen nach Auffassung der Behörde über den Offenbarungsgehalt in der ursprünglich eingereichten Fassung hinausgehen (Regel 70.2(c)):

4. Etwaige zusätzliche Bemerkungen:

V. Begründete Feststellung nach Artikel 35(2) hinsichtlich der Neuheit, der erfinderischen Tätigkeit und der gewerblichen Anwendbarkeit; Unterlagen und Erklärungen zur Stützung dieser Feststellung

1. Feststellung

Neuheit (N)	Ja: Ansprüche	1-8
	Nein: Ansprüche	
Erfinderische Tätigkeit (ET)	Ja: Ansprüche	1-8
	Nein: Ansprüche	
Gewerbliche Anwendbarkeit (GA)	Ja: Ansprüche	1-8
	Nein: Ansprüche	

2. Unterlagen und Erklärungen

siehe Beiblatt

zu Punkt V:

1. Die Erfindung betrifft ein Verfahren zur Erkennung eines vorgegebenen Wortschatzes in gesprochener Sprache durch einen Rechner.

Das Dokument D1= TAKAMI ET AL.: 'A successive state splitting algorithm for efficient allophone modeling' INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP) 1992, Bd. 1, 23. - 26. März 1992, Seiten 573-576, SAN FRANCISCO (USA) (= nächster Stand der Technik) offenbart ein Verfahren zur Spracherkennung, bei dem in einem Markov-Modell Zustände aufgespalten werden, sodaß die Wahrscheinlichkeitsdichtefunktionen der jeweiligen Zustände ebenfalls aufgespalten sind. Die Erfinder erkennen die technische Aufgabe, daß im Rahmen einer Sprecheradaption ein Spracherkennungsverfahren nach dem Stand der Technik zu einer verminderten Güte führen kann. Zur Lösung der technischen Aufgabe schlagen die Erfinder vor, ein statistisches Kriterium einzuführen, anhand dessen entschieden wird, ob der zu erkennende Wortschatz durch das Hinzufügen weiterer Zustände verbessert modelliert werden kann; bei Erfüllung des statistischen Kriteriums werden bereits existierende Zustände aufgespalten.

Der Gegenstand des Anspruchs 1 ist neu. Da die Lösung der technischen Aufgabe nicht naheliegend ist, beinhaltet Anspruch 1 ebenfalls einen erfinderischen Schritt.

2. Der Anspruch 7 beansprucht eine Anordnung zur Erkennung eines vorgegebenen Wortschatzes in gesprochener Sprache mit einer Prozessoreinheit. Da Vorrichtungsanspruch 7 dem Verfahrensanspruch 1 entspricht, ist Anspruch 7 ebenfalls neu und erfinderisch.

Bemerkungen:

3. Aufgrund der PCT-Verfahrensprozedur erscheinen die geänderten Vorrichtungsansprüche 2 und 8 das Kriterium der Einheitlichkeit der Erfindung nicht zu erfüllen (Regel 13 PCT). Die "Veränderung des Wortschatzes zur Laufzeit des Verfahrens" wird als ein "besonderes technisches Merkmal" gewertet, das zu einer zweiten erfinderischen Idee führt.

Sprecheradaption durch eine MAP-Schätzung der Hidden-Markov-Modell-Parameter durchgeführt.

Generell verschlechtern sich Ergebnisse eines Verfahrens zur
5 Erkennung gesprochener Sprache, sobald charakteristische
Merkmale der gesprochenen Sprache von charakteristischen
Merkmalen der Trainingsdaten abweichen. Beispiele für
charakteristische Merkmale sind Sprechereigenschaften oder
akustische Kontexte, die sich in Form von Verschleifungen auf
10 die Artikulation der Phoneme auswirken.

Der in [2] verfolgte Ansatz zur Sprecheradaption besteht
darin, Parameterwerte der Hidden-Markov-Modelle
"nachzuschätzen", wobei diese nach Verarbeitung "offline",
15 d.h. nicht zur Laufzeit des Verfahrens zur Spracherkennung,
durchgeführt wird.

In J. Takami et al. "Successive State Splitting Algorithm for
Efficient Allophone Modeling", ICASSP 1992, März 1992, Seiten
20 573 bis 576, San Francisco, USA, wird ein Verfahren zur
Erkennung eines vorgegebenen Wortschatzes in gesprochener
Sprache offenbart, bei dem in einem Hidden-Markov-Modell
Zustände aufgespalten werden. Dafür wird auch die
Wahrscheinlichkeitsdichtefunktion der jeweiligen Zustände
25 aufgespalten.

Die **Aufgabe** der Erfindung besteht darin, eine Anordnung und
ein Verfahren zur Erkennung eines vorgegebenen Wortschatzes
in gesprochener Sprache anzugeben, wobei insbesondere eine
30 Anpassung des akustischen Modells zur Laufzeit (also
"Online") vollzogen wird.

Diese Aufgabe wird gemäß den Merkmalen der unabhängigen
Patentansprüche gelöst.

35 Zur Lösung der Aufgabe wird ein Verfahren zur Erkennung eines
vorgegebenen Wortschatzes in gesprochener Sprache durch einen

11 07.07.00

4a

Rechner angegeben, in dem aus der gesprochenen Sprache ein Sprachsignal bestimmt wird. Das Sprachsignal wird einer Signalanalyse unterworfen, woraus Merkmalsvektoren zur Beschreibung des digitalisierten Sprachsignals hervorgehen.

- 5 Eine globale Suche wird zur Abbildung der Merkmalsvektoren auf eine in modellierter Form vorliegende Sprache durchgeführt, wobei jedes Phonem der Sprache durch ein modifiziertes Hidden-Markov-Modell und jeder Zustand des modifizierten Hidden-Markov-Modells durch eine

Patentansprüche

1. Verfahren zur Erkennung eines vorgegebenen Wortschatzes
in gesprochener Sprache durch einen Rechner,
- 5 a) bei dem aus der gesprochenen Sprache ein
digitalisiertes Sprachsignal bestimmt wird,
- b) bei dem auf dem digitalisierten Sprachsignal eine
Signalanalyse durchgeführt wird, woraus
Merkmalsvektoren zur Beschreibung des digitalisierten
10 Sprachsignals hervorgehen,
- c) bei dem eine globale Suche zur Abbildung der
Merkmalsvektoren auf eine in modellierter Form
vorliegende Sprache durchgeführt wird, wobei Phoneme
der Sprache durch ein modifiziertes Hidden-Markov-
15 Modell und jeder Zustand des Hidden-Markov-Modells
durch eine Wahrscheinlichkeitsdichtefunktion
beschrieben wird,
- d) bei dem die Wahrscheinlichkeitsdichtefunktion durch
Veränderung des Wortschatzes angepaßt wird, indem die
20 Wahrscheinlichkeitsdichtefunktion in eine erste
Wahrscheinlichkeitsdichtefunktion und in eine zweite
Wahrscheinlichkeitsdichtefunktion aufgespalten wird,
falls der Abfall eines Entropiewertes unterhalb einer
vorgegebenen Schranke liegt,
- 25 e) bei dem von der globalen Suche eine erkannte Wortfolge
bereitgestellt wird.
2. Verfahren zur Erkennung eines vorgegebenen Wortschatzes
in gesprochener Sprache durch einen Rechner, insbesondere
30 nach Anspruch 1,
- a) bei dem aus der gesprochenen Sprache ein
digitalisiertes Sprachsignal bestimmt wird,
- b) bei dem auf dem digitalisierten Sprachsignal eine
Signalanalyse durchgeführt wird, woraus
35 Merkmalsvektoren zur Beschreibung des digitalisierten
Sprachsignals hervorgehen,

- 5 c) bei dem eine globale Suche zur Abbildung der Merkmalsvektoren auf eine in modellierter Form vorliegende Sprache durchgeführt wird, wobei Phoneme der Sprache durch ein modifiziertes Hidden-Markov-Modell und jeder Zustand des Hidden-Markov-Modells durch eine Wahrscheinlichkeitsdichtefunktion beschrieben wird,
- 10 d) bei dem die Wahrscheinlichkeitsdichtefunktion durch Veränderung des Wortschatzes angepaßt wird, indem die Wahrscheinlichkeitsdichtefunktion in eine erste Wahrscheinlichkeitsdichtefunktion und in eine zweite Wahrscheinlichkeitsdichtefunktion aufgespalten wird,
- 15 e) bei dem die Veränderung des Wortschatzes zur Laufzeit des Verfahrens durchgeführt wird,
- f) bei dem von der globalen Suche eine erkannte Wortfolge bereitgestellt wird.
3. Verfahren nach Anspruch 1 oder 2, bei dem die Veränderung des Wortschatzes bedingt ist durch Hinzufügen eines Wortes zum Wortschatz oder bei dem sich Aussprachegewohnheiten eines Sprechers ändern.
- 20 4. Verfahren nach einem der vorhergehenden Ansprüche, bei dem die erste Wahrscheinlichkeitsdichtefunktion und die zweite Wahrscheinlichkeitsdichtefunktion jeweils mindestens eine Gaußverteilung umfassen.
- 25 5. Verfahren nach Anspruch 4, bei dem für die erste Wahrscheinlichkeitsdichtefunktion und für die zweite Wahrscheinlichkeitsdichtefunktion gleiche Standardabweichungen, ein erster Mittelwert der ersten Wahrscheinlichkeitsdichtefunktion und ein zweiter Mittelwert der zweiten Wahrscheinlichkeitsdichtefunktion ermittelt werden, wobei der erste Mittelwert von dem
- 30 zweiten Mittelwert verschieden ist.
- 35

6. Verfahren nach einem der vorhergehenden Ansprüche,
bei dem die Aufspaltung mehrfach durchgeführt wird.
7. Anordnung zur Erkennung eines vorgegebenen Wortschatzes
in gesprochener Sprache mit einer Prozessoreinheit, die
derart eingerichtet ist, daß
- a) aus der gesprochenen Sprache ein digitalisiertes
Sprachsignal bestimmbar ist,
 - b) auf dem digitalisierten Sprachsignal eine
Signalanalyse durchführbar ist, woraus
Merkmalsvektoren zur Beschreibung des digitalisierten
Sprachsignals hervorgehen,
 - c) eine globale Suche zur Abbildung der Merkmalsvektoren
auf eine in modellierter Form vorliegende Sprache
erfolgt, wobei Phoneme der Sprache durch ein
modifiziertes Hidden-Markov-Modell und jeder Zustand
des Hidden-Markov-Modells durch eine
Wahrscheinlichkeitsdichtefunktion beschreibbar ist,
 - d) die Wahrscheinlichkeitsdichtefunktion durch
Veränderung des Wortschatzes angepaßt wird, indem die
Wahrscheinlichkeitsdichtefunktion in eine erste
Wahrscheinlichkeitsdichtefunktion und in eine zweite
Wahrscheinlichkeitsdichtefunktion aufgespalten wird,
falls der Abfall eines Entropiewertes unterhalb einer
vorgegebenen Schranke liegt,
 - e) von der globalen Suche eine erkannte Wortfolge
bereitgestellt wird.
8. Anordnung zur Erkennung eines vorgegebenen Wortschatzes
in gesprochener Sprache mit einer Prozessoreinheit, die
derart eingerichtet ist, daß
- a) aus der gesprochenen Sprache ein digitalisiertes
Sprachsignal bestimmbar ist,
 - b) auf dem digitalisierten Sprachsignal eine
Signalanalyse durchführbar ist, woraus
Merkmalsvektoren zur Beschreibung des digitalisierten
Sprachsignals hervorgehen,

11 07 07 00

- 5 c) eine globale Suche zur Abbildung der Merkmalsvektoren
auf eine in modellierter Form vorliegende Sprache
erfolgt, wobei Phoneme der Sprache durch ein
modifiziertes Hidden-Markov-Modell und jeder Zustand
des Hidden-Markov-Modells durch eine
Wahrscheinlichkeitsdichtefunktion beschreibbar ist,
- 10 d) die Wahrscheinlichkeitsdichtefunktion durch
Veränderung des Wortschatzes angepaßt wird, indem die
Wahrscheinlichkeitsdichtefunktion in eine erste
Wahrscheinlichkeitsdichtefunktion und in eine zweite
Wahrscheinlichkeitsdichtefunktion aufgespalten wird,
- e) die Veränderung des Wortschatzes zur Laufzeit des
Verfahrens durchgeführt wird,
- 15 f) von der globalen Suche eine erkannte Wortfolge
bereitgestellt wird.

GEÄNDERTES BLATT

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

RECEIVED

FEB - 9 2001

TC 2600 MAILROOM

09700143

2644
Translation

Applicant's or agent's file reference GR 98P1687P	FOR FURTHER ACTION See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)	
International application No. PCT/DE99/01323	International filing date (day/month/year) 03 May 1999 (03.05.99)	Priority date (day/month/year) 11 May 1998 (11.05.98)
International Patent Classification (IPC) or national classification and IPC G10L 5/06		
Applicant SIEMENS AKTIENGESELLSCHAFT		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 4 sheets, including this cover sheet.

☒ This report is also accompanied by ANNEXES, i.e., sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consist of a total of 6 sheets.

3. This report contains indications relating to the following items:

- I ☒ Basis of the report
- II ☐ Priority
- III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- IV ☐ Lack of unity of invention
- V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI ☐ Certain documents cited
- VII ☐ Certain defects in the international application
- VIII ☐ Certain observations on the international application

Date of submission of the demand 02 November 1999 (02.11.99)	Date of completion of this report 09 August 2000 (09.08.2000)
Name and mailing address of the IPEA/EP	Authorized officer
Facsimile No.	Telephone No.

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/DE99/01323

I. Basis of the report

1. This report has been drawn on the basis of (Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.):

- ☐ the international application as originally filed.
- ☒ the description, pages 1-3,5-14, as originally filed,
 pages _____, filed with the demand,
 pages 4,4a, filed with the letter of 03 July 2000 (03.07.2000),
 pages _____, filed with the letter of _____.
- ☒ the claims, Nos. _____, as originally filed,
 Nos. _____, as amended under Article 19,
 Nos. _____, filed with the demand,
 Nos. 1-8, filed with the letter of 03 July 2000 (03.07.2000),
 Nos. _____, filed with the letter of _____.
- ☒ the drawings, sheets/fig 1/1, as originally filed,
 sheets/fig _____, filed with the demand,
 sheets/fig _____, filed with the letter of _____,
 sheets/fig _____, filed with the letter of _____.

2. The amendments have resulted in the cancellation of:

- ☐ the description, pages _____
- ☐ the claims, Nos. _____
- ☐ the drawings, sheets/fig _____

3. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed, as indicated in the Supplemental Box (Rule 70.2(c)).

4. Additional observations, if necessary:

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No.

PCT/DE 99/01323

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Claims	1-8	YES
	Claims		NO
Inventive step (IS)	Claims	1-8	YES
	Claims		NO
Industrial applicability (IA)	Claims	1-8	YES
	Claims		NO

2. Citations and explanations

1. The invention concerns a method for computer recognition of a predefined vocabulary in spoken language.

Document D1, TAKAMI ET AL.: 'A successive state splitting algorithm for efficient allophone modelling', INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), 1992, Vol. 1, 23-26 March 1992, pages 573-576, SAN FRANCISCO (USA), considered the closest prior art, discloses a method for voice recognition in which, in a Markov model, states are split such that the probability density functions of the individual states are likewise split. The inventors identify the technical problem that, in the context of speaker adaptation, a voice recognition method according to the prior art can lead to a reduction in quality. To solve the technical problem, the inventors propose that a statistical criterion be introduced on the basis of which it is decided whether the modelling of the vocabulary for recognition can be improved by the addition of further states; when the statistical criterion is satisfied, already existing states are split.

The subject matter of Claim 1 is novel. Since the solution of the technical problem is not obvious, Claim 1 also involves an inventive step.

2. Claim 7 concerns a device for recognising a predefined vocabulary in spoken language using a processor unit. Since device Claim 7 corresponds to method Claim 1, Claim 7 is also novel and inventive.

Comments:

3. In the PCT proceedings, the amended device Claims 2 and 8 do not appear to satisfy the criterion of unity of invention (PCT Rule 13). "Altering the vocabulary during the method" is deemed a "special technical feature", which leads to a second inventive concept.

According to [2], thus, it is recognized that a speaker-dependent system for speech recognition normally supplies better results than a speaker-independent system, insofar as adequate training data are available that enable a modelling of the speaker-dependent system. However, the speaker-independent system achieves the better results as soon as the set of speaker-specific training data is limited. One possibility for performance enhancement of both systems, i.e. of both the speaker-dependent as well as the speaker-independent system for speech recognition, is comprised in employing previously stored datasets of a plurality of speakers such that a small set of training data also suffices for modelling a new speaker with adequate quality. Such a training method is called speaker adaptation. In [2], the speaker adaptation is particularly implemented by a MAP estimate of the hidden Markov model parameters.

Results of a method for recognizing spoken language generally deteriorate as soon as characteristic features of the spoken language deviate from characteristic features of the training data. Examples of characteristic features are speaker qualities or acoustic features that influence the articulation of the phonemes in the form of slurring.

The approach embarked upon in [2] for speaker adaptation is comprised in "post-estimating" parameter values of the hidden Markov models, whereby this processing is implemented "offline", i.e. not at the run time of the method for speech recognition.

The **object** of the invention is comprised in specifying an arrangement and a method for recognizing a predetermined vocabulary in spoken language, whereby, in particular, an adaptation of the acoustic model is accomplished at the run time (i.e., "online").

This object is achieved according to the features of the independent patent claims.

For achieving the object, a method for recognizing a predetermined vocabulary in spoken language with a computer is recited wherein a voice signal is determined from the spoken language. The voice signal is subjected to a signal analysis from which feature vectors for describing the digitalized voice signal proceed. A global search is implemented for imaging the feature vectors onto a

Patent Claims

1. Method for recognizing a predetermined vocabulary in spoken language with a computer,

a) whereby a digitalized voice signal is determined from the spoken language;

b) whereby a signal analysis is implemented on the digitalized voice signal, feature vectors for describing the digitalized voice signal proceeding therefrom;

c) whereby a global search ensues for imaging the feature vectors onto a language present in modelled form, whereby phoneme [sic] of the language can be described by a modified hidden Markov model and each status of the hidden Markov model can be described by a probability density function;

d) whereby the probability density function is adapted by modification of the vocabulary in that the probability density function is split into a first probability density function and into a second probability density function;

e) whereby the global search offers a recognized word sequence.

2. Method according to claim 1, whereby the probability density function is split into the first and into the second probability density function if the drop of an entropy value lies below a predetermined threshold.

3. Method according to claim 1 or 2, whereby the modification of the vocabulary is implemented at the run time of the method.

4. Method according to one of the preceding claims, whereby the modification of the vocabulary is caused by the addition of a word to the vocabulary or whereby pronunciation habits of a speaker change.

5. Method according to one of the preceding claims, whereby the first probability density function and the second probability density function respectively comprise at least one Gaussian distribution.

6. Method according to claim 5, whereby identical standard deviations, a first average of the first probability density function and a second average of the second probability density function are determined for the first probability density

function and for the second probability density function, whereby the first average differs from the second average.

7. Method according to one of the preceding claims, whereby the splitting is multiply implemented.

5 8. Arrangement for recognizing a predetermined vocabulary in spoken language comprising a processor unit that is configured such that

- a) a digitalized voice signal can be determined from the spoken language;
- b) a signal analysis can be implemented on the digitalized voice signal, feature vectors for describing the digitalized voice signal proceeding
10 therefrom;
- c) a global search ensues for imaging the feature vectors onto a language present in modelled form, whereby phoneme [sic] of the language can be described by a modified hidden Markov model and each status of the hidden Markov model can be described by a probability density function;
- 15 d) the probability density function is adapted by modification of the vocabulary in that the probability density function is split into a first probability density function and into a second probability density function;
- e) the global search offers a recognized word sequence.

PTO/PCT Rec'd 09 NOV 2000

Beschreibung**Anordnung und Verfahren zur Erkennung eines vorgegebenen Wortschatzes in gesprochener Sprache durch einen Rechner**

5

Die Erfindung betrifft eine Anordnung und ein Verfahren zur Erkennung eines vorgegebenen Wortschatzes in gesprochener Sprache durch einen Rechner.

- 10 Ein Verfahren und eine Anordnung zur Erkennung gesprochener Sprache sind aus [1] bekannt. Bei der Erkennung gesprochener Sprache werden, insbesondere bis zum Erhalt einer erkannten Wortfolge aus einem digitalisierten Sprachsignal, eine
15 Signalanalyse und eine globale Suche, die auf ein akustisches Modell und ein linguistisches Modell der zu erkennenden Sprache zurückgreift, durchgeführt. Ein akustisches Modell basiert auf einem Phoneminventar, das anhand von Hidden-Markov-Modellen (HMMs) realisiert ist. Während der globalen Suche werden für Merkmalsvektoren, die aus der Signalanalyse
20 hervorgegangen sind, mit Hilfe des akustischen Modells eine passende Wortfolge ermittelt und diese als erkannte Wortfolge ausgegeben. Die zu erkennenden Wörter sind in einem Aussprachelexikon zusammen mit einer phonetischen Umschrift abgespeichert. Der Zusammenhang ist ausführlich in [1]
25 dargestellt.

Zur Erläuterung der nachfolgenden Ausführungen wird an dieser Stelle kurz auf die verwendeten Begriffe eingegangen.

- 30 Die Signalanalyse als Phase der computerbasierten Spracherkennung umfaßt insbesondere eine Fouriertransformation des digitalisierten Sprachsignals und eine sich daran anschließende Merkmalsextraktion. Aus [1] geht hervor, daß die Signalanalyse alle zehn Millisekunden
35 erfolgt. Aus sich überlappenden Zeitabschnitten mit einer Dauer von z.B. jeweils 25 Millisekunden werden anhand der Signalanalyse ungefähr 30 Merkmale ermittelt und zu einem

Merkmalsvektor zusammengefaßt. Die Komponenten des Merkmalsvektors beschreiben die spektrale Energieverteilung des zugehörigen Signalausschnitts. Um diese Energieverteilung zu erhalten, wird auf jedem Signalabschnitt (25ms-
5 Zeitabschnitt) eine Fouriertransformation durchgeführt. Aus der Darstellung des Signals im Frequenzbereich resultieren die Komponenten des Merkmalsvektors. Nach der Signalanalyse liegt das digitalisierte Sprachsignal in Form von Merkmalsvektoren vor.

10

Diese Merkmalsvektoren werden der globalen Suche, einer weiteren Phase der Spracherkennung, zugeführt. Wie bereits erwähnt, bedient sich die globale Suche des akustischen Modells und ggf. des linguistischen Modells, um die Folge von
15 Merkmalsvektoren auf Einzelteile der als Modell vorliegenden Sprache (Vokabular) abzubilden. Eine Sprache setzt sich aus einer vorgegebenen Anzahl vom Lauten, sog. Phonemen, zusammen, deren Gesamtheit als Phoneminventar bezeichnet wird. Das Vokabular wird durch Phonemfolgen modelliert und in
20 einem Aussprachelexikon abgespeichert. Jedes Phonem wird durch mindestens ein HMM modelliert. Mehrere HMMs ergeben einen stochastischen Automaten, der Zustände und Zustandsübergänge (Transitionen) umfaßt. Mit HMMs läßt sich der zeitliche Ablauf des Auftretens bestimmter
25 Merkmalsvektoren (selbst innerhalb eines Phonems) modellieren. Ein entsprechendes Phonem-Modell umfaßt dabei eine vorgegebene Anzahl von Zuständen, die linear hintereinander angeordnet sind. Ein Zustand eines HMMs stellt einen Teil eines Phonems (bspw. mit einer Dauer von 10ms)
30 dar. Jeder Zustand ist verknüpft mit einer Emissionswahrscheinlichkeit, die insbesondere nach Gauß verteilt ist, für die Merkmalsvektoren und mit Transitionswahrscheinlichkeiten für die möglichen Übergänge. Mit der Emissionsverteilung wird einem Merkmalsvektor eine
35 Wahrscheinlichkeit zugeordnet, mit der dieser Merkmalsvektor in einem zugehörigen Zustand beobachtet wird. Die möglichen Übergänge sind ein direkter Übergang von einem Zustand in

einen nächsten Zustand, ein Wiederholen des Zustands und ein Überspringen des Zustands.

Eine Aneinanderreihung von HMM-Zustände mit den zugehörigen
5 Übergängen über die Zeit wird als Trellis bezeichnet. Um die
akustische Wahrscheinlichkeit eines Wortes zu bestimmen,
verwendet man insbesondere das Prinzip der dynamischen
Programmierung: Es wird der Pfad durch die Trellis gesucht,
der den geringsten Fehler aufweist bzw. der durch die größte
10 Wahrscheinlichkeit für ein zu erkennendes Wort bestimmt ist.

Das Ergebnis der globalen Suche ist die Ausgabe bzw.
Bereitstellung einer erkannten Wortfolge, die sich unter
Berücksichtigung des akustischen Modells (Phoneminventar) für
15 jedes einzelne Wort und des Sprachmodells für die Abfolge von
Wörtern ergibt.

Aus [2] ist ein Verfahren zur Sprecheradaption, basierend auf
einer MAP-Schätzung (MAP = maximum a posteriori) von HMM-
20 Parametern bekannt.

So ist es laut [2] anerkannt, daß ein sprecherabhängiges
System zur Spracherkennung normalerweise bessere Ergebnisse
als ein sprecherunabhängiges System liefert, sofern
ausreichend Trainingsdaten verfügbar sind, die eine
25 Modellierung des sprecherabhängigen Systems ermöglichen.
Sobald jedoch die Menge der sprecherspezifischen
Trainingsdaten beschränkt ist, erreicht das
sprecherunabhängige System die besseren Resultate. Eine
Möglichkeit zur Leistungssteigerung beider Systeme, also
30 sowohl des sprecherabhängigen als auch des
sprecherunabhängigen Systems zur Spracherkennung, besteht
darin, die vorab gespeicherten Datensätze mehrerer Sprecher,
derart zu benutzen, daß auch eine kleine Menge Trainingsdaten
ausreicht, um einen neuen Sprecher in ausreichender Qualität
35 zu modellieren. Solch ein Trainingsverfahren wird
Sprecheradaption genannt. In [2] wird insbesondere die

Sprecheradaption durch eine MAP-Schätzung der Hidden-Markov-Modell-Parameter durchgeführt.

5 Generell verschlechtern sich Ergebnisse eines Verfahrens zur
Erkennung gesprochener Sprache, sobald charakteristische
Merkmale der gesprochenen Sprache von charakteristischen
Merkmalen der Trainingsdaten abweichen. Beispiele für
charakteristische Merkmale sind Sprechereigenschaften oder
akustische Kontexte, die sich in Form von Verschleifungen auf
10 die Artikulation der Phoneme auswirken.

Der in [2] verfolgte Ansatz zur Sprecheradaption besteht
darin, Parameterwerte der Hidden-Markov-Modelle
"nachzuschätzen", wobei diese nach Verarbeitung "offline",
15 d.h. nicht zur Laufzeit des Verfahrens zur Spracherkennung,
durchgeführt wird.

Die **Aufgabe** der Erfindung besteht darin, eine Anordnung und
ein Verfahren zur Erkennung eines vorgegebenen Wortschatzes
20 in gesprochener Sprache anzugeben, wobei insbesondere eine
Anpassung des akustischen Modells zur Laufzeit (also
"Online") vollzogen wird.

Diese Aufgabe wird gemäß den Merkmalen der unabhängigen
25 Patentansprüche gelöst.

Zur Lösung der Aufgabe wird ein Verfahren zur Erkennung eines
vorgegebenen Wortschatzes in gesprochener Sprache durch einen
Rechner angegeben, in dem aus der gesprochenen Sprache ein
30 Sprachsignal bestimmt wird. Das Sprachsignal wird einer
Signalanalyse unterworfen, woraus Merkmalsvektoren zur
Beschreibung des digitalisierten Sprachsignals hervorgehen.
Eine globale Suche wird zur Abbildung der Merkmalsvektoren
auf eine in modellierter Form vorliegende Sprache
35 durchgeführt, wobei jedes Phonem der Sprache durch ein
modifiziertes Hidden-Markov-Modell und jeder Zustand des
modifizierten Hidden-Markov-Modells durch eine

Wahrscheinlichkeitsdichtefunktion beschrieben wird. Es erfolgt eine Anpassung der Wahrscheinlichkeitsdichtefunktion derart, daß sie in eine erste Wahrscheinlichkeitsdichtefunktion und in eine zweite Wahrscheinlichkeitsdichtefunktion aufgespalten wird. Schließlich wird von der globalen Suche eine erkannte Wortfolge bereitgestellt.

Hierbei sei angemerkt, daß die Wahrscheinlichkeitsdichtefunktion, die in eine erste und in eine zweite Wahrscheinlichkeitsdichtefunktion aufgespalten wird, eine Emissionsverteilung für einen vorgegebenen Zustand des modifizierten Hidden-Markov-Modells darstellen kann, wobei diese Emissionsverteilung auch eine Überlagerung mehrerer Wahrscheinlichkeitsdichtefunktionen, z.B. Gauß-Kurven (Gauß'sche Wahrscheinlichkeitsdichteverteilungen), enthalten kann.

Eine erkannte Wortfolge kann dabei auch einzelne Lauten bzw. nur ein einzelnes Wort umfassen.

Sollte im Rahmen der globalen Suche eine Erkennung mit einem hohen Wert für den Abstand zwischen gesprochener Sprache und von der globalen Suche ermittelten dazugehöriger Wortfolge behaftet sein, so kann die Zuordnung eines Nullwortes erfolgen, welches Nullwort anzeigt, das die gesprochene Sprache nicht mit ausreichender Güte erkannt wird.

Es ist ein Vorteil der Erfindung, durch die Aufspaltung der Wahrscheinlichkeitsdichtefunktion in einem durch die Merkmalsvektoren aufgespannten Merkmalsraum neue Bereiche zu schaffen, die signifikante Information in Bezug auf die zu erkennenden digitalisierten Sprachdaten aufweisen und damit eine verbesserte Erkennung zu gewährleisten.

Eine Ausgestaltung besteht darin, daß die Wahrscheinlichkeitsdichtefunktion in die erste und in die zweite Wahrscheinlichkeitsdichtefunktion aufgespalten wird,

falls der Abfall eines Entropiewertes unterhalb einer vorgegebenen Schranke liegt.

Die Aufspaltung der Wahrscheinlichkeitsdichtefunktion in
5 Abhängigkeit von einem Entropiewert erweist sich in der Praxis als äußerst vorteilhaft.

Die Entropie ist allgemein ein Maß für eine Unsicherheit bei einer Vorhersage eines statistischen Ereignisses. Die
10 Entropie ist insbesondere mathematisch bestimmbar für Gauß-Verteilungen, wobei eine direkte logarithmische Abhängigkeit zwischen der Streuung σ und der Entropie besteht.

Eine andere Ausgestaltung der Erfindung besteht darin, daß
15 die Wahrscheinlichkeitsdichtefunktionen, insbesondere die erste und die zweite Wahrscheinlichkeitsdichtefunktion jeweils mindestens eine Gauß-Verteilung umfassen.

Die Wahrscheinlichkeitsdichtefunktion des Zustandes wird
20 durch eine Summe mehrerer Gaußverteilungen angenähert. Die einzelnen Gaußverteilungen werden Moden genannt. Bei dem vorgestellten Verfahren werden die Moden insbesondere isoliert voneinander betrachtet. Bei jedem einzelnen Aufspaltvorgang wird eine Mode in zwei Moden aufgeteilt. Wenn
25 die Wahrscheinlichkeitsdichtefunktion aus M Moden gebildet wurde, so wird sie nach dem Aufspaltvorgang aus M+1 Moden gebildet. Wird eine Mode beispielsweise als eine Gaußverteilung angenommen, so kann eine Entropie berechnet werden, wie im Ausführungsbeispiel gezeigt wird.

30

Eine Online-Adaption ist deshalb vorteilhaft, weil das Verfahren nach wie vor Sprache erkennt, ohne in einer gesonderten Trainingsphase auf die Veränderung des Wortschatzes eingestellt werden zu müssen. Es erfolgt eine
35 Selbstadaption, die insbesondere notwendig wird durch eine veränderte Koartikulation der Sprecher aufgrund eines Hinzufügens eines neuen Wortes.

Die Online-Adaption erfordert demnach keine gesonderte Berechnung der Wahrscheinlichkeitsdichtefunktionen, die wiederum für eine Nicht-Verfügbarkeit des Systems zur
5 Spracherkennung verantwortlich wäre.

Eine Weiterbildung der Erfindung besteht darin, daß für die erste Wahrscheinlichkeitsdichtefunktion und für die zweite Wahrscheinlichkeitsdichtefunktion gleiche
10 Standardabweichungen bestimmt werden. Ein erster Mittelwert der ersten Wahrscheinlichkeitsdichtefunktion und ein zweiter Mittelwert der zweiten Wahrscheinlichkeitsdichtefunktion werden derart bestimmt, daß der erste Mittelwert von dem zweiten Mittelwert verschieden ist.

15 Dies ist ein Beispiel für die Gewichtung der aus der Wahrscheinlichkeitsdichtefunktion aufgespaltenen ersten und zweiten Wahrscheinlichkeitsdichtefunktion. Es sind auch beliebig andere Gewichtungen vorstellbar, die auf den
20 jeweiligen Anwendungsfall anzupassen sind.

Schließlich ist es eine Weiterbildung, daß das Verfahren mehrfach hintereinander durchgeführt wird und somit eine wiederholte Aufspaltung der Wahrscheinlichkeitsdichtefunktion
25 erfolgt.

Weiterbildungen der Erfindung ergeben sich aus den abhängigen Ansprüchen.

30 Eine andere Lösung der Aufgabe besteht darin, eine Anordnung mit einer Prozessoreinheit anzugeben, welche Prozessoreinheit derart eingerichtet ist, daß folgende Schritte durchführbar sind:

35 a) aus der gesprochenen Sprache wird ein digitalisiertes Sprachsignal bestimmt;;

- b) auf dem digitalisierten Sprachsignal erfolgt eine Signalanalyse, woraus Merkmalsvektoren zur Beschreibung des digitalisierten Sprachsignals hervorgehen;
- 5 c) eine globale Suche zur Abbildung der Merkmalsvektoren erfolgt auf eine in modellierter Form vorliegende Sprache, wobei Phoneme der Sprache durch ein modifiziertes Hidden-Markov-Modell und jeder Zustand des Hidden-Markov-Modells durch eine
- 10 Wahrscheinlichkeitsdichtefunktion beschreibbar ist;
- d) die wird Wahrscheinlichkeitsdichtefunktion durch Veränderung des Wortschatzes angepaßt, indem die Wahrscheinlichkeitsdichtefunktion in eine erste Wahrscheinlichkeitsdichtefunktion und in eine zweite
- 15 Wahrscheinlichkeitsdichtefunktion aufgespalten wird;
- e) von der globalen Suche wird eine erkannte Wortfolge bereitgestellt.

Diese Anordnung ist insbesondere geeignet zur Durchführung

20 des erfindungsgemäßen Verfahrens oder einer seiner vorstehend erläuterten Weiterbildungen.

Ausführungsbeispiele der Erfindung werden nachfolgend anhand der Zeichnung dargestellt und erläutert.

25

Es zeigt

Fig.1 eine Anordnung bzw. ein Verfahren zur Erkennung

30 gesprochener Sprache.

30

In Figur 1 sind eine Anordnung bzw. ein Verfahren zur Erkennung gesprochener Sprache dargestellt. Zur Erläuterung der nachstehend verwendeten Begriffe sei auf die Beschreibungseinleitung verwiesen.

35

Ein digitalisiertes Sprachsignal 101 wird in einer Signalanalyse 102 einer Fouriertransformation 103 mit

anschließender Merkmalsextraktion 104 unterzogen. Die Merkmalsvektoren 105 werden an ein System zur globalen Suche 106 übermittelt. Die globale Suche 106 berücksichtigt sowohl ein akustisches Modell 107 als auch ein linguistisches Modell 108 zur Bestimmung der erkannten Wortfolge 109. Aus dem digitalisierten Sprachsignal 101 geht somit die erkannte Wortfolge 109 hervor.

In dem akustischen Modell 107 wird das Phoneminventar anhand von Hidden-Markov-Modellen nachgebildet.

Eine Wahrscheinlichkeitsdichtefunktion eines Zustands des Hidden-Markov-Modells wird durch eine Aufsummierung einzelner Gaußscher Moden angenähert. Eine Mode ist insbesondere eine Gaußglocke. Durch Aufsummierung mehrerer Moden entsteht eine Mischung einzelner Gaußglocken und damit eine Modellierung der Emissionswahrscheinlichkeitsdichtefunktion. Anhand eines statistischen Kriteriums wird entschieden, ob der zu erkennende Wortschatz des Spracherkenners durch das Hinzufügen weiterer Moden verbessert modelliert werden kann. Im Fall der vorliegenden Erfindung wird dies insbesondere bei Erfüllung des statistischen Kriteriums durch inkrementelles Aufspalten bereits existierender Moden erreicht.

Die Entropie ist definiert durch

$$H_p = - \int_{-\infty}^{\infty} p(\bar{x}) \log_2 p(\bar{x}) d\bar{x} \quad (1)$$

unter der Annahme, daß $p(\bar{x})$ eine Gauß-Verteilung mit einer diagonalen Kovarianzmatrix ist, also

$$p(\bar{x}) = \mathcal{N}(\bar{\mu}, \sigma_n) = \frac{1}{\sqrt{(2\pi)^N}} \frac{1}{\prod_n \sigma_n} \cdot \exp \left(- \frac{1}{2} \sum_n \frac{(x_n - \mu_n)^2}{\sigma_n^2} \right) \quad (2)$$

erhält man

$$H_p = \sum_{n=1}^N \log_2 \sqrt{2\pi e} \sigma_n \quad (3),$$

5 wobei

μ den Erwartungswert,
 σ_n die Streuung für jede Komponente n und
 N die Dimension des Merkmalsraums

bezeichnen.

10

Die wahre Verteilung $p(\vec{x})$ ist nicht bekannt. Sie wird insbesondere als Gaußverteilung angenommen. Im akustischen Modell wird die Wahrscheinlichkeit $p(\vec{x})$ anhand von Stichproben angenähert mit

15

$$\hat{p}(\vec{x}) = \mathcal{N}(\vec{\mu}, \sigma_n),$$

wobei

20

$$\vec{\mu} = \frac{1}{L} \sum_{l=1}^L \vec{x}_l$$

einen Mittelwert über L Beobachtungen darstellt. Die korrespondierende Entropie als Funktion von $\hat{\mu}$ ist gegeben durch

25

$$H_{\hat{p}}(\hat{\mu}) = - \int_{-\infty}^{\infty} p(\vec{x}) \log_2 \hat{p}(\vec{x}) d\vec{x} \quad (4),$$

was schließlich zu

30

$$H_{\hat{p}}(\hat{\mu}) = H_p + \sum_{n=1}^N \frac{(\mu_n - \hat{\mu}_n)^2}{\sigma_n^2} \log_2 \sqrt{e} \quad (5)$$

führt.

Der Erwartungswert $E\{(\mu_n - \hat{\mu}_n)^2\}$ beträgt $\frac{1}{L} \sigma_n^2$, so daß der Erwartungswert von $H_{\hat{p}}(\hat{\mu})$ gegeben ist als

5

$$H_{\hat{p}} = E\{H_{\hat{p}}(\hat{\mu})\} = H_p + \frac{N}{L} \log_2 \sqrt{e} \quad (6).$$

Für die Entropie einer Mode, die mit einer Gauß-Verteilung mit einer diagonalen Kovarianzmatrix bestimmt wird, ergibt sich also Gleichung (3). Der Prozeß wird nun mit einer Schätzung angenähert. Die Entropie des angenäherten Prozesses ergibt sich zu

$$\hat{H} = H + \frac{N}{L} \log_2 \sqrt{e} \quad (7).$$

15

Je größer die Anzahl L der Stichproben ist, um so besser wird die Abschätzung und um so mehr nähert sich die geschätzte Entropie \hat{H} der wahren Entropie H an.

20 Es soll nun

$$p(\bar{x}) = \mathcal{N}(\bar{\mu}, \sigma_n) \quad (8)$$

die aufzuteilende Mode sein. Ferner wird angenommen, daß die zwei Gauß-Verteilungen, die durch den Aufteilungsprozeß entstehen, identische Standardabweichungen σ^S haben und gleich gewichtet sind. Dies ergibt

$$\hat{p}^S(\bar{x}) = \frac{1}{2} \mathcal{N}(\bar{\mu}_1^S, \sigma^S) + \frac{1}{2} \mathcal{N}(\bar{\mu}_2^S, \sigma^S) \quad (9).$$

30

Unter der Annahme, daß $\mu_1 \approx \hat{\mu}_1$, $\mu_2 \approx \hat{\mu}_2$ und daß μ_1 ausreichend weit entfernt von μ_2 ist, ergibt sich die Entropie der aufgespaltenen Wahrscheinlichkeitsdichtefunktion jeweils zu

$$\hat{H}^S = 1 - \sum_{n=1}^N \log_2 \sqrt{2\pi e} \sigma_n^S + \frac{1}{2} \left(\log_2 \sqrt{e} \frac{N}{L_1} + \log_2 \sqrt{e} \frac{N}{L_2} \right) \quad (10).$$

Als Aufteilungskriterium wird eine Verminderung der Entropie
5 durch den Aufspaltungsvorgang gefordert, also

$$\hat{H} - \hat{H}^S > C \quad (11),$$

wobei C (mit $C > 0$) eine Konstante ist, die den gewünschten
10 Abfall der Entropie darstellt. Wird

$$\frac{L}{2} = L_1 = L_2 \quad (12)$$

angenommen, so ergibt sich hierdurch
15

$$\sum_{n=1}^N \log_2 \frac{\sigma_n}{\sigma_n^S} > \log_2 \sqrt{e} \frac{N}{L} + 1 + C \quad (13).$$

Eine Möglichkeit, die Lage der Mittelpunkte der beiden neuen
Moden zu bestimmen, wird im folgenden aufgezeigt. Eine
20 bevorzugte Vorgabe besteht darin, daß Kriterium zum
Aufspalten zu erfüllen. In dem angeführten Beispiel wird $\hat{\mu}_1^S$
der Wert von $\hat{\mu}$ zugewiesen. $\hat{\mu}_2^S$ erhält eine Maximum-
Likelihood-Schätzung derjenigen Beobachtungen, die im
Viterbi-Pfad auf $\hat{\mu}$ abgebildet werden. Diese Bestimmungen
25 zeigen lediglich eine Möglichkeit auf, ohne daß eine
Einschränkung des vorgestellten Verfahrens auf diese
Möglichkeit beabsichtigt ist.

Die folgenden Schritte der Beispielanwendung zeigen die
30 Einbettung in eine Anordnung zur Spracherkennung bzw. ein
Verfahren zur Spracherkennung.

Schritt 1: Initialisierung: $\bar{\mu}_1^S = \bar{\mu}$, $\bar{\mu}_2^S = \bar{\mu}$.

Schritt 2: Erkennen der Äußerung , Analysieren des Viterbi-Pfads;

5 Schritt 3: Für jeden Zustand und für jede Mode des Viterbi-Pfades:

Schritt 3.1: Bestimme σ_n ;

10 Schritt 3.2: Bestimme L_2 auf Grundlage derjenigen Beobachtungen, die näher an $\bar{\mu}_2^S$ als an $\bar{\mu}_1^S$ liegen und setze $L = L_2$. Falls $\bar{\mu}_2^S$ und $\bar{\mu}_1^S$ identisch sind, so ordne die zweite Hälfte der Merkmalsvektoren $\bar{\mu}_2^S$ und die
15 erste Hälfte der Merkmalsvektoren $\bar{\mu}_1^S$ zu.

Schritt 3.3: Bestimme σ_n^S entsprechend auf Grundlage der L_2 -Äußerungen;

20 Schritt 3.4: Ermittle $\bar{\mu}_2^S$ neu auf Grundlage des Mittelwerts derjenigen Beobachtungen, die näher an $\bar{\mu}_2^S$ als an $\bar{\mu}_1^S$ liegen;

25 Schritt 3.5: Werte Aufteilungskriterium nach Gleichung (13) aus;

Schritt 3.6: Falls Aufteilungskriterium nach Gleichung (13) positiv ist, generiere zwei neue Moden mit den Mittelpunkten $\bar{\mu}_1^S$
30 und $\bar{\mu}_2^S$.

Schritt 4: Gehe zu Schritt 2. . .

Im Rahmen dieses Dokuments wurden folgende Veröffentlichungen zitiert:

[1] N. Haberland et al.: "Sprachunterricht - Wie funktioniert die computerbasierte Spracherkennung?", c't - Magazin für Computertechnik - 5/1998, Heinz Heise Verlag, Hannover, 1998, Seiten 120 bis 125.

[2] C. H. Lee et al.: "Speaker Adaptation Based on MAP Estimation of HMM Parameters"; Proc. IEEE Intern. Conference on Acoustics, Speech and Signal Processing, Seiten II-588 bis II-561.

Patentansprüche

1. Verfahren zur Erkennung eines vorgegebenen Wortschatzes
in gesprochener Sprache durch einen Rechner,
 - a) bei dem aus der gesprochenen Sprache ein
digitalisiertes Sprachsignal bestimmt wird,
 - b) bei dem auf dem digitalisierten Sprachsignal eine
Signalanalyse durchgeführt wird, woraus
Merkmalsvektoren zur Beschreibung des digitalisierten
Sprachsignals hervorgehen,
 - c) bei dem eine globale Suche zur Abbildung der
Merkmalsvektoren auf eine in modellierter Form
vorliegende Sprache durchgeführt wird, wobei Phoneme
der Sprache durch ein modifiziertes Hidden-Markov-
Modell und jeder Zustand des Hidden-Markov-Modells
durch eine Wahrscheinlichkeitsdichtefunktion
beschrieben wird,
 - d) bei dem die Wahrscheinlichkeitsdichtefunktion durch
Veränderung des Wortschatzes angepaßt wird, indem die
Wahrscheinlichkeitsdichtefunktion in eine erste
Wahrscheinlichkeitsdichtefunktion und in eine zweite
Wahrscheinlichkeitsdichtefunktion aufgespalten wird,
 - e) bei dem von der globalen Suche eine erkannte Wortfolge
bereitgestellt wird.
2. Verfahren nach Anspruch 1,
bei dem die Wahrscheinlichkeitsdichtefunktion in die
erste und in die zweite Wahrscheinlichkeitsdichtefunktion
aufgespalten wird, falls der Abfall eines Entropiewertes
unterhalb einer vorgegebenen Schranke liegt.
3. Verfahren nach Anspruch 1 oder 2,
bei dem die Veränderung des Wortschatzes zur Laufzeit des
Verfahrens durchgeführt wird.
4. Verfahren nach einem der vorhergehenden Ansprüche,
bei dem die Veränderung des Wortschatzes bedingt ist

durch Hinzufügen eines Wortes zum Wortschatz oder bei dem sich Aussprachegewohnheiten eines Sprechers ändern.

5. Verfahren nach einem der vorhergehenden Ansprüche,
5 bei dem die erste Wahrscheinlichkeitsdichtefunktion und die zweite Wahrscheinlichkeitsdichtefunktion jeweils mindestens eine Gaußverteilung umfassen.
6. Verfahren nach Anspruch 5,
10 bei dem für die erste Wahrscheinlichkeitsdichtefunktion und für die zweite Wahrscheinlichkeitsdichtefunktion gleiche Standardabweichungen, ein erster Mittelwert der ersten Wahrscheinlichkeitsdichtefunktion und ein zweiter Mittelwert der zweiten Wahrscheinlichkeitsdichtefunktion
15 ermittelt werden, wobei der erste Mittelwert von dem zweiten Mittelwert verschieden ist.
7. Verfahren nach einem der vorhergehenden Ansprüche,
20 bei dem die Aufspaltung mehrfach durchgeführt wird.
8. Anordnung zur Erkennung eines vorgegebenen Wortschatzes in gesprochener Sprache mit einer Prozessoreinheit, die derart eingerichtet ist, daß
25
 - a) aus der gesprochenen Sprache ein digitalisiertes Sprachsignal bestimmbar ist,
 - b) auf dem digitalisierten Sprachsignal eine Signalanalyse durchführbar ist, woraus Merkmalsvektoren zur Beschreibung des digitalisierten Sprachsignals hervorgehen,
30
 - c) eine globale Suche zur Abbildung der Merkmalsvektoren auf eine in modellierter Form vorliegende Sprache erfolgt, wobei Phoneme der Sprache durch ein modifiziertes Hidden-Markov-Modell und jeder Zustand des Hidden-Markov-Modells durch eine
35 Wahrscheinlichkeitsdichtefunktion beschreibbar ist,

- 5
- d) die Wahrscheinlichkeitsdichtefunktion durch Veränderung des Wortschatzes angepaßt wird, indem die Wahrscheinlichkeitsdichtefunktion in eine erste Wahrscheinlichkeitsdichtefunktion und in eine zweite Wahrscheinlichkeitsdichtefunktion aufgespalten wird,
 - e) von der globalen Suche eine erkannte Wortfolge bereitgestellt wird.

Zusammenfassung

Anordnung und Verfahren zur Erkennung eines vorgegebenen Wortschatzes in gesprochener Sprache durch einen Rechner

5

Bei der Spracherkennung werden Phoneme einer Sprache durch ein Hidden-Markov-Modell modelliert, wobei jeder Zustand des Hidden-Markov-Modells durch eine

10

Wahrscheinlichkeitsdichtefunktion beschrieben wird. Zur Spracherkennung eines veränderten Wortschatzes wird die Wahrscheinlichkeitsdichtefunktion in eine erste und eine zweite Wahrscheinlichkeitsdichtefunktion aufgespalten.

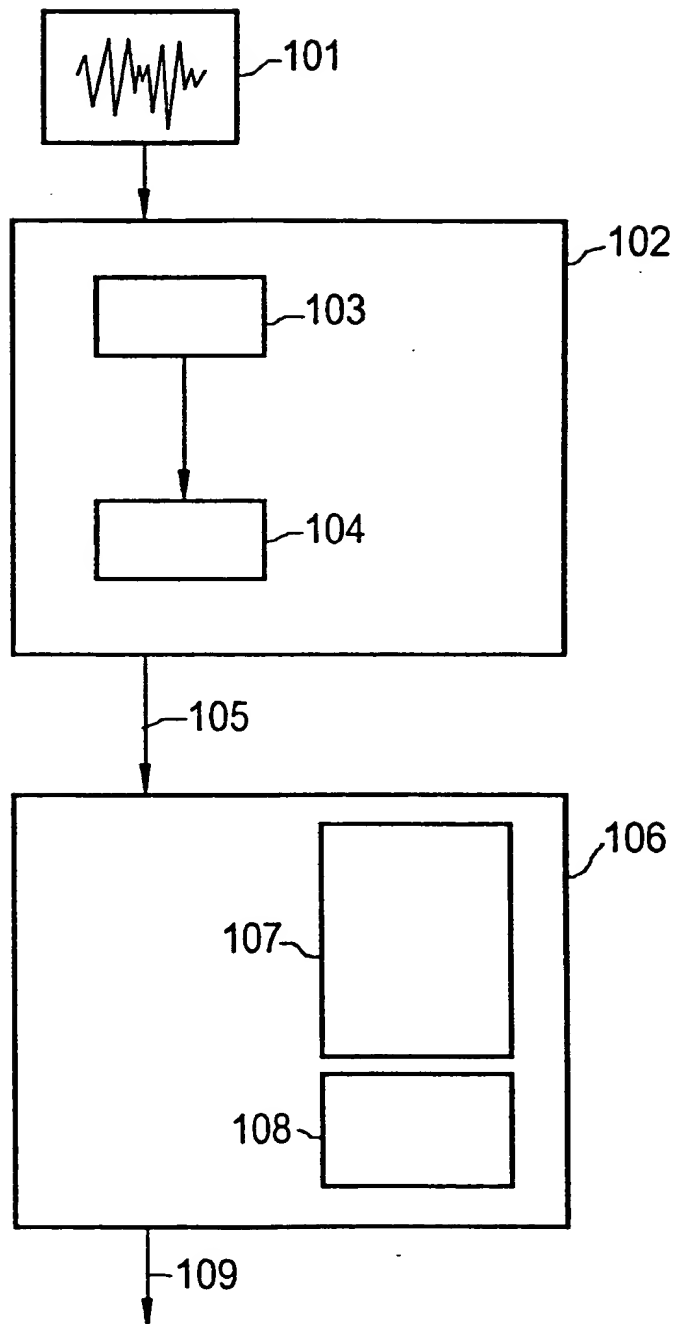
Dadurch wird es möglich, Veränderungen der Sprachgewohnheit eines Sprechers zu kompensieren oder ein neues Wort dem

15

Wortschatz des Spracherkenners hinzuzufügen und dabei sicherzustellen, daß dieses neue Wort mit ausreichender Güte von den bereits im Spracherkenner vorhandenen Wörtern unterschieden und somit erkannt wird.

20

Figur 1



VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT
AUF DEM GEBIET DES PATENTWESENS

Absender: INTERNATIONALE RECHERCHENBEHÖRDE

PCT/DE 99/01323

19 NOV 2000

PCT

An
SIEMENS AG
Postfach 22 16 34
80506 München
GERMANY

MITTEILUNG ÜBER DIE ÜBERMITTLUNG DES
INTERNATIONALEN RECHERCHENBERICHTS
ODER DER ERKLÄRUNG

(Regel 44.1 PCT)

ZT nach P/Ri

Eing. 21. Okt. 1999

GR

Absendedatum
(Tag/Monat/Jahr)

19/10/1999

Aktenzeichen des Anmelders oder Anwalts

GR-98P1688P GR 98P1687P

WEITERES VORGEHEN

siehe Punkte 1 und 4 unten

Internationales Aktenzeichen

PCT/DE 99/01323

Internationales Anmeldedatum

(Tag/Monat/Jahr)

03/05/1999

Anmelder

SIEMENS AKTIENGESellschaft et al.

1. ☒ Dem Anmelder wird mitgeteilt, daß der internationale Recherchenbericht erstellt wurde und ihm hiermit übermittelt wird.

Einreichung von Änderungen und einer Erklärung nach Artikel 19:

Der Anmelder kann auf eigenen Wunsch die Ansprüche der internationalen Anmeldung ändern (siehe Regel 46):

Bis wann sind Änderungen einzureichen?

Die Frist zur Einreichung solcher Änderungen beträgt üblicherweise zwei Monate ab der Übermittlung des internationalen Recherchenberichts; weitere Einzelheiten sind den Anmerkungen auf dem Beiblatt zu entnehmen.

Wo sind Änderungen einzureichen?

Unmittelbar beim Internationalen Büro der WIPO, 34, CHEMIN des Colombettes, CH-1211 Genf 20,
Telefaxnr.: (41-22) 740.14.35

Nähere Hinweise sind den Anmerkungen auf dem Beiblatt zu entnehmen.

2. ☐ Dem Anmelder wird mitgeteilt, daß kein internationaler Recherchenbericht erstellt wird und daß ihm hiermit die Erklärung nach Artikel 17(2a) übermittelt wird.

3. ☐ Hinsichtlich des Widerspruchs gegen die Entrichtung einer zusätzlichen Gebühr (zusätzlicher Gebühren) nach Regel 40.2 wird dem Anmelder mitgeteilt, daß

☐ der Widerspruch und die Entscheidung hierüber zusammen mit seinem Antrag auf Übermittlung des Wortlauts sowohl des Widerspruchs als auch der Entscheidung hierüber an die Bestimmungsämter dem Internationalen Büro übermittelt worden sind.

☐ noch keine Entscheidung über den Widerspruch vorliegt; der Anmelder wird benachrichtigt, sobald eine Entscheidung getroffen wurde.

4. **Weiteres Vorgehen:** Der Anmelder wird auf folgendes aufmerksam gemacht:

Kurz nach Ablauf von 18 Monaten seit dem Prioritätsdatum wird die internationale Anmeldung vom Internationalen Büro veröffentlicht. Will der Anmelder die Veröffentlichung verhindern oder auf einen späteren Zeitpunkt verschieben, so muß gemäß Regel 90 bis bzw. 90 bis 3 vor Abschluß der technischen Vorbereitungen für die internationale Veröffentlichung eine Erklärung über die Zurücknahme der internationalen Anmeldung oder des Prioritätsanspruchs beim Internationalen Büro eingehen.

Innerhalb von 19 Monaten seit dem Prioritätsdatum ist ein Antrag auf internationale vorläufige Prüfung einzureichen, wenn der Anmelder den Eintritt in die nationale Phase bis zu 30 Monaten seit dem Prioritätsdatum (in manchen Ämtern sogar noch länger) verschieben möchte.

Innerhalb von 20 Monaten seit dem Prioritätsdatum muß der Anmelder die für den Eintritt in die nationale Phase vorgeschriebenen Handlungen vor allen Bestimmungsämtern vornehmen, die nicht innerhalb von 19 Monaten seit dem Prioritätsdatum in der Anmeldung oder einer nachträglichen Auswahlerklärung ausgewählt wurden oder nicht ausgewählt werden konnten, da für sie Kapitel II des Vertrages nicht verbindlich ist.

Name und Postanschrift der Internationalen Recherchenbehörde



Europäisches Patentamt, P.B. 5818 Patentaan 2
NL-2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Ahmed Soliman

Diese Anmerkungen sollen grundlegende Hinweise zur Einreichung von Änderungen gemäß Artikel 19 geben. Diesen Anmerkungen liegen die Erfordernisse des Vertrags über die internationale Zusammenarbeit auf dem Gebiet des Patentwesens (PCT), der Ausführungsordnung und der Verwaltungsrichtlinien zu diesem Vertrag zugrunde. Bei Abweichungen zwischen diesen Anmerkungen und obengenannten Texten sind letztere maßgebend. Nähere Einzelheiten sind dem PCT-Leitfaden für Anmelder, einer Veröffentlichung der WIPO, zu entnehmen.

Die in diesen Anmerkungen verwendeten Begriffe "Artikel", "Regel" und "Abschnitt" beziehen sich jeweils auf die Bestimmungen des PCT-Vertrags, der PCT-Ausführungsordnung bzw. der PCT-Verwaltungsrichtlinien.

HINWEISE ZU ÄNDERUNGEN GEMÄSS ARTIKEL 19

Nach Erhalt des internationalen Recherchenberichts hat der Anmelder die Möglichkeit, einmal die Ansprüche der internationalen Anmeldung zu ändern. Es ist jedoch zu betonen, daß, da alle Teile der internationalen Anmeldung (Ansprüche, Beschreibung und Zeichnungen) während des internationalen vorläufigen Prüfungsverfahrens geändert werden können, normalerweise keine Notwendigkeit besteht, Änderungen der Ansprüche nach Artikel 19 einzureichen, außer wenn der Anmelder z.B. zum Zwecke eines vorläufigen Schutzes die Veröffentlichung dieser Ansprüche wünscht oder ein anderer Grund für eine Änderung der Ansprüche vor ihrer internationalen Veröffentlichung vorliegt. Weiterhin ist zu beachten, daß ein vorläufiger Schutz nur in einigen Staaten erhältlich ist.

Welche Teile der internationalen Anmeldung können geändert werden?

Im Rahmen von Artikel 19 können nur die Ansprüche geändert werden.

In der internationalen Phase können die Ansprüche auch nach Artikel 34 vor der mit der internationalen vorläufigen Prüfung beauftragten Behörde geändert (oder nochmals geändert) werden. Die Beschreibung und die Zeichnungen können nur nach Artikel 34 vor der mit der internationalen vorläufigen Prüfung beauftragten Behörde geändert werden.

Beim Eintritt in die nationale Phase können alle Teile der internationalen Anmeldung nach Artikel 28 oder gegebenenfalls Artikel 41 geändert werden.

Bis wann sind Änderungen einzureichen?

Innerhalb von zwei Monaten ab der Übermittlung des internationalen Recherchenberichts oder innerhalb von sechzehn Monaten ab dem Prioritätsdatum, je nachdem, welche Frist später abläuft. Die Änderungen gelten jedoch als rechtzeitig eingereicht, wenn sie dem Internationalen Büro nach Ablauf der maßgebenden Frist, aber noch vor Abschluß der technischen Vorbereitungen für die internationale Veröffentlichung (Regel 46.1) zugehen.

Wo sind die Änderungen nicht einzureichen?

Die Änderungen können nur beim Internationalen Büro, nicht aber beim Anmeldeamt oder der Internationalen Recherchenbehörde eingereicht werden (Regel 46.2).

Falls ein Antrag auf internationale vorläufige Prüfung eingereicht wurde/wird, siehe unten.

In welcher Form können Änderungen erfolgen?

Eine Änderung kann erfolgen durch Streichung eines oder mehrerer ganzer Ansprüche, durch Hinzufügung eines oder mehrerer neuer Ansprüche oder durch Änderung des Wortlauts eines oder mehrerer Ansprüche in der eingereichten Fassung.

Für jedes Anspruchsblatt, das sich aufgrund einer oder mehrerer Änderungen von dem ursprünglich eingereichten Blatt unterscheidet, ist ein Ersatzblatt einzureichen.

Alle Ansprüche, die auf einem Ersatzblatt erscheinen, sind mit arabischen Ziffern zu numerieren. Wird ein Anspruch gestrichen, so brauchen die anderen Ansprüche nicht neu numeriert zu werden. Im Fall einer Neunummerierung sind die Ansprüche fortlaufend zu numerieren (Verwaltungsrichtlinien, Abschnitt 205 b)).

Die Änderungen sind in der Sprache abzufassen, in der die internationale Anmeldung veröffentlicht wird.

Welche Unterlagen sind den Änderungen beizufügen?

Begleitschreiben (Abschnitt 205 b)):

Die Änderungen sind mit einem Begleitschreiben einzureichen.

Das Begleitschreiben wird nicht zusammen mit der internationalen Anmeldung und den geänderten Ansprüchen veröffentlicht. Es ist nicht zu verwechseln mit der "Erklärung nach Artikel 19(1)" (siehe unten, "Erklärung nach Artikel 19 (1)").

Das Begleitschreiben ist nach Wahl des Anmelders in englischer oder französischer Sprache abzufassen. Bei englischsprachigen internationalen Anmeldungen ist das Begleitschreiben aber ebenfalls in englischer, bei französischsprachigen internationalen Anmeldungen in französischer Sprache abzufassen.

ANMERKUNGEN ZU FORMBLATT PCT/ISA/220 (Übersetzung)

Im Begleitschreiben sind die Unterschiede zwischen den Ansprüchen in der eingereichten Fassung und den geänderten Ansprüchen anzugeben. So ist insbesondere zu jedem Anspruch in der internationalen Anmeldung anzugeben (gleichlautende Angaben zu verschiedenen Ansprüchen können zusammengefaßt werden), ob

- i) der Anspruch unverändert ist;
- ii) der Anspruch gestrichen worden ist;
- iii) der Anspruch neu ist;
- iv) der Anspruch einen oder mehrere Ansprüche in der eingereichten Fassung ersetzt;
- v) der Anspruch auf die Teilung eines Anspruchs in der eingereichten Fassung zurückzuführen ist.

Im folgenden sind Beispiele angegeben, wie Änderungen im Begleitschreiben zu erläutern sind:

1. [Wenn anstelle von ursprünglich 48 Ansprüchen nach der Änderung einiger Ansprüche 51 Ansprüche existieren]:
"Die Ansprüche 1 bis 29, 31, 32, 34, 35, 37 bis 48 werden durch geänderte Ansprüche gleicher Numerierung ersetzt; Ansprüche 30, 33 und 36 unverändert; neue Ansprüche 49 bis 51 hinzugefügt."
2. [Wenn anstelle von ursprünglich 15 Ansprüchen nach der Änderung aller Ansprüche 11 Ansprüche existieren]:
"Geänderte Ansprüche 1 bis 11 treten an die Stelle der Ansprüche 1 bis 15."
3. [Wenn ursprünglich 14 Ansprüche existierten und die Änderungen darin bestehen, daß einige Ansprüche gestrichen werden und neue Ansprüche hinzugefügt werden]:
Ansprüche 1 bis 6 und 14 unverändert; Ansprüche 7 bis 13 gestrichen; neue Ansprüche 15, 16 und 17 hinzugefügt. "Oder" Ansprüche 7 bis 13 gestrichen; neue Ansprüche 15, 16 und 17 hinzugefügt; alle übrigen Ansprüche unverändert."
4. [Wenn verschiedene Arten von Änderungen durchgeführt werden]:
"Ansprüche 1-10 unverändert; Ansprüche 11 bis 13, 18 und 19 gestrichen; Ansprüche 14, 15 und 16 durch geänderten Anspruch 14 ersetzt; Anspruch 17 in geänderte Ansprüche 15, 16 und 17 unterteilt; neue Ansprüche 20 und 21 hinzugefügt."

"Erklärung nach Artikel 19(1)" (Regel 46.4)

Den Änderungen kann eine Erklärung beigefügt werden, mit der die Änderungen erläutert und ihre Auswirkungen auf die Beschreibung und die Zeichnungen dargelegt werden (die nicht nach Artikel 19 (1) geändert werden können).

Die Erklärung wird zusammen mit der internationalen Anmeldung und den geänderten Ansprüchen veröffentlicht.

Sie ist in der Sprache abzufassen, in der die internationale Anmeldung veröffentlicht wird.

Sie muß kurz gehalten sein und darf, wenn in englischer Sprache abgefaßt oder ins Englische übersetzt, nicht mehr als 500 Wörter umfassen.

Die Erklärung ist nicht zu verwechseln mit dem Begleitschreiben, das auf die Unterschiede zwischen den Ansprüchen in der eingereichten Fassung und den geänderten Ansprüchen hinweist, und ersetzt letzteres nicht. Sie ist auf einem gesonderten Blatt einzureichen und in der Überschrift als solche zu kennzeichnen, vorzugsweise mit den Worten "Erklärung nach Artikel 19 (1)".

Die Erklärung darf keine herabsetzenden Äußerungen über den internationalen Recherchenbericht oder die Bedeutung von in dem Bericht angeführten Veröffentlichungen enthalten. Sie darf auf im internationalen Recherchenbericht angeführte Veröffentlichungen, die sich auf einen bestimmten Anspruch beziehen, nur im Zusammenhang mit einer Änderung dieses Anspruchs Bezug nehmen.

Auswirkungen eines bereits gestellten Antrags auf internationale vorläufige Prüfung

Ist zum Zeitpunkt der Einreichung von Änderungen nach Artikel 19 bereits ein Antrag auf internationale vorläufige Prüfung gestellt worden, so sollte der Anmelder in seinem Interesse gleichzeitig mit der Einreichung der Änderungen beim Internationalen Büro auch eine Kopie der Änderungen bei der mit der internationalen vorläufigen Prüfung beauftragten Behörde einreichen (siehe Regel 62.2 a), erster Satz).

Auswirkungen von Änderungen hinsichtlich der Übersetzung der internationalen Anmeldung beim Eintritt in die nationale Phase

Der Anmelder wird darauf hingewiesen, daß bei Eintritt in die nationale Phase möglicherweise anstatt oder zusätzlich zu der Übersetzung der Ansprüche in der eingereichten Fassung eine Übersetzung der nach Artikel 19 geänderten Ansprüche an die bestimmten/ausgewählten Ämter zu übermitteln ist.

Nähere Einzelheiten über die Erfordernisse jedes bestimmten/ausgewählten Amtes sind Band II des PCT-Leitfadens für Anmelder zu entnehmen.

56

**VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT
AUF DEM GEBIET DES PATENTWESENS**

PCT

INTERNATIONALER RECHERCHENBERICHT

(Artikel 18 sowie Regeln 43 und 44 PCT)

Aktenzeichen des Anmelders oder Anwalts GR 98P1688P	WEITERES VORGEHEN siehe Mitteilung über die Übermittlung des internationalen Recherchenberichts (Formblatt PCT/ISA/220) sowie, soweit zutreffend, nachstehender Punkt 5	
Internationales Aktenzeichen PCT/DE 99/ 01323	Internationales Anmeldedatum (Tag/Monat/Jahr) 03/05/1999	(Früheste) Prioritätsdatum (Tag/Monat/Jahr) 11/05/1998
Anmelder SIEMENS AKTIENGESELLSCHAFT et al.		

Dieser internationale Recherchenbericht wurde von der internationalen Recherchenbehörde erstellt und wird dem Anmelder gemäß Artikel 18 übermittelt. Eine Kopie wird dem internationalen Büro übermittelt.

Dieser internationale Recherchenbericht umfaßt insgesamt 3 Blätter.



Darüber hinaus liegt ihm jeweils eine Kopie der in diesem Bericht genannten Unterlagen zum Stand der Technik bei.

1. Grundlage des Berichts

- a. Hinsichtlich der Sprache ist die internationale Recherche auf der Grundlage der internationalen Anmeldung in der Sprache durchgeführt worden, in der sie eingereicht wurde, sofern unter diesem Punkt nichts anderes angegeben ist.



Die internationale Recherche ist auf der Grundlage einer bei der Behörde eingereichten Übersetzung der internationalen Anmeldung (Regel 23.1 b)) durchgeführt worden.

- b. Hinsichtlich der in der internationalen Anmeldung offenbaren Nucleotid- und/oder Aminosäuresequenz ist die internationale Recherche auf der Grundlage des Sequenzprotokolls durchgeführt worden, das



in der internationalen Anmeldung in schriftlicher Form enthalten ist.



zusammen mit der internationalen Anmeldung in computerlesbarer Form eingereicht worden ist.



bei der Behörde nachträglich in schriftlicher Form eingereicht worden ist.



bei der Behörde nachträglich in computerlesbarer Form eingereicht worden ist.



Die Erklärung, daß das nachträglich eingereichte schriftliche Sequenzprotokoll nicht über den Offenbarungsgehalt der internationalen Anmeldung im Anmeldezeitpunkt hinausgeht, wurde vorgelegt.



Die Erklärung, daß die in computerlesbarer Form erfaßten Informationen dem schriftlichen Sequenzprotokoll entsprechen, wurde vorgelegt.

2.



Bestimmte Ansprüche haben sich als nicht recherchierbar erwiesen (siehe Feld I).

3.



Mangelnde Einheitlichkeit der Erfindung (siehe Feld II).

4. Hinsichtlich der Bezeichnung der Erfindung



wird der vom Anmelder eingereichte Wortlaut genehmigt.



wurde der Wortlaut von der Behörde wie folgt festgesetzt:

5. Hinsichtlich der Zusammenfassung



wird der vom Anmelder eingereichte Wortlaut genehmigt.



wurde der Wortlaut nach Regel 38.2b) in der in Feld III angegebenen Fassung von der Behörde festgesetzt. Der Anmelder kann der Behörde innerhalb eines Monats nach dem Datum der Absendung dieses internationalen Recherchenberichts eine Stellungnahme vorlegen.

6. Folgende Abbildung der Zeichnungen ist mit der Zusammenfassung zu veröffentlichen: Abb. Nr. 1



wie vom Anmelder vorgeschlagen



keine der Abb.



weil der Anmelder selbst keine Abbildung vorgeschlagen hat.



weil diese Abbildung die Erfindung besser kennzeichnet.

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
 IPK 6 G10L5/06 G10L3/00

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchiertes Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)
 IPK 6 G10L

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der Internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	JP 09 152886 A (ATR INTERPRETING TELECOM) 10. Juni 1997 (1997-06-10) Zusammenfassung	1,8
P,A	-& US 5 839 105 A (OSTENDORF ET AL.) 17. November 1998 (1998-11-17) Spalte 4 -Spalte 6, Zeile 31	1,8
X	TAKAMI ET AL.: "A successive state splitting algorithm for efficient allophone modeling" INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP) 1992, Bd. 1, 23. - 26. März 1992, Seiten 573-576, XP000341211 SAN FRANCISCO, CA, US ISBN: 0-7803-0532-9 Seite 574 -Seite 575, linke Spalte -/-	1,8

☒ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

☒ Siehe Anhang Patentfamilie

* Besondere Kategorien von angegebenen Veröffentlichungen :

"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfindertischer Tätigkeit beruhend betrachtet werden

"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfindertischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der Internationalen Recherche

12. Oktober 1999

Abmeldedatum des Internationalen Recherchenberichts

19/10/1999

Name und Postanschrift der Internationalen Recherchenbehörde
 Europäisches Patentamt, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
 Fax: (+31-70) 340-3018

Bevollmächtigter Bediensteter

Lange, J

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	LEE ET AL.: "Speaker adaptation based on map estimation of HMM parameters" INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP) 1993, Bd. 2, 27. - 30. April 1993, Seiten 558-561, XP000427850 MINNEAPOLIS, MN, US ISBN: 0-7803-0946-4 in der Anmeldung erwähnt das ganze Dokument	1,8

INTERNATIONAL SEARCH REPORT

In addition on patent family members

International Application No

PCT/DE 99/01323

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
JP 09152886 A	10-06-1997	JP 2871561 B	17-03-1999
		US 5839105 A	17-11-1998



<p>(51) Internationale Patentklassifikation ⁶ : G10L 5/06, 3/00</p>	A2	<p>(11) Internationale Veröffentlichungsnummer: WO 99/59135</p> <p>(43) Internationales Veröffentlichungsdatum: 18. November 1999 (18.11.99)</p>		
<table style="width: 100%; border: none;"> <tr> <td style="width: 50%; vertical-align: top; padding: 5px;"> <p>(21) Internationales Aktenzeichen: PCT/DE99/01323</p> <p>(22) Internationales Anmeldedatum: 3. Mai 1999 (03.05.99)</p> <p>(30) Prioritätsdaten: 198 21 057.4 11. Mai 1998 (11.05.98) DE</p> <p>(71) Anmelder (für alle Bestimmungsstaaten ausser US): SIEMENS AKTIENGESELLSCHAFT (DE/DE); Wittelsbacherplatz 2, D-80333 München (DE).</p> <p>(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): BUB, Udo (DE/DE); Klarweinstrasse 18, D-81247 München (DE). HÖGE, Harald (DE/DE); Obertaxetweg 6B, D-82131 Gauting (DE).</p> <p>(74) Gemeinsamer Vertreter: SIEMENS AKTIENGESELLSCHAFT; Postfach 22 16 34, D-80506 München (DE).</p> </td> <td style="width: 50%; vertical-align: top; padding: 5px;"> <p>(81) Bestimmungsstaaten: US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Veröffentlicht <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p> </td> </tr> </table>			<p>(21) Internationales Aktenzeichen: PCT/DE99/01323</p> <p>(22) Internationales Anmeldedatum: 3. Mai 1999 (03.05.99)</p> <p>(30) Prioritätsdaten: 198 21 057.4 11. Mai 1998 (11.05.98) DE</p> <p>(71) Anmelder (für alle Bestimmungsstaaten ausser US): SIEMENS AKTIENGESELLSCHAFT (DE/DE); Wittelsbacherplatz 2, D-80333 München (DE).</p> <p>(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): BUB, Udo (DE/DE); Klarweinstrasse 18, D-81247 München (DE). HÖGE, Harald (DE/DE); Obertaxetweg 6B, D-82131 Gauting (DE).</p> <p>(74) Gemeinsamer Vertreter: SIEMENS AKTIENGESELLSCHAFT; Postfach 22 16 34, D-80506 München (DE).</p>	<p>(81) Bestimmungsstaaten: US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Veröffentlicht <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p>
<p>(21) Internationales Aktenzeichen: PCT/DE99/01323</p> <p>(22) Internationales Anmeldedatum: 3. Mai 1999 (03.05.99)</p> <p>(30) Prioritätsdaten: 198 21 057.4 11. Mai 1998 (11.05.98) DE</p> <p>(71) Anmelder (für alle Bestimmungsstaaten ausser US): SIEMENS AKTIENGESELLSCHAFT (DE/DE); Wittelsbacherplatz 2, D-80333 München (DE).</p> <p>(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): BUB, Udo (DE/DE); Klarweinstrasse 18, D-81247 München (DE). HÖGE, Harald (DE/DE); Obertaxetweg 6B, D-82131 Gauting (DE).</p> <p>(74) Gemeinsamer Vertreter: SIEMENS AKTIENGESELLSCHAFT; Postfach 22 16 34, D-80506 München (DE).</p>	<p>(81) Bestimmungsstaaten: US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Veröffentlicht <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p>			
<p>(54) Title: ARRANGEMENT AND METHOD FOR COMPUTER RECOGNITION OF A PREDEFINED VOCABULARY IN SPOKEN LANGUAGE</p> <p>(54) Bezeichnung: ANORDNUNG UND VERFAHREN ZUR ERKENNUNG EINES VORGEGEBENEN WORTSCHATZES IN GESPROCHENER SPRACHE DURCH EINEN RECHNER</p> <p>(57) Abstract</p> <p>In speech recognition, phonemes of a language are modelled, whereby each state of the hidden Markov model is described by means of a probability density function. The probability density function is divided into a first and a second probability function in order to recognize modifications of vocabulary. This makes it possible to compensate for changes in the spoken habits of a speaker, enables a new word to be added to the vocabulary of the speech recognition system and guarantees that the new word can be recognized and distinguished from the words that are already present in the speech recognition system with a sufficient amount of quality.</p> <p>(57) Zusammenfassung</p> <p>Bei der Spracherkennung werden Phoneme einer Sprache durch ein Hidden-Markov-Modell modelliert, wobei jeder Zustand des Hidden-Markov-Modells durch eine Wahrscheinlichkeitsdichtefunktion beschrieben wird. Zur Spracherkennung eines veränderten Wortschatzes wird die Wahrscheinlichkeitsdichtefunktion in eine erste und eine zweite Wahrscheinlichkeitsdichtefunktion aufgespalten. Dadurch wird es möglich, Veränderungen der Sprachgewohnheit eines Sprechers zu kompensieren oder ein neues Wort dem Wortschatz des Spracherkenners hinzuzufügen und dabei sicherzustellen, daß dieses neue Wort mit ausreichender Güte von den bereits im Spracherkenner vorhandenen Wörtern unterschieden und somit erkannt wird.</p>				
<pre> graph TD 101[101] --> 102[102] subgraph 102 103[103] --> 104[104] end 102 -- 105 --> 106[106] subgraph 106 107[107] --> 108[108] end 106 -- 109 --> 109[109] </pre>				

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidshan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland		Republik Mazedonien	TR	Türkei
BG	Bulgarien	HU	Ungarn	ML	Mali	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MN	Mongolei	UA	Ukraine
BR	Brasilien	IL	Israel	MR	Mauretanien	UG	Uganda
BY	Belarus	IS	Island	MW	Malawi	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	MX	Mexiko	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NE	Niger	VN	Vietnam
CG	Kongo	KE	Kenia	NL	Niederlande	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NO	Norwegen	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	NZ	Neuseeland		
CM	Kamerun			PL	Polen		
CN	China	KR	Republik Korea	PT	Portugal		
CU	Kuba	KZ	Kasachstan	RO	Rumänien		
CZ	Tschechische Republik	LC	St. Lucia	RU	Russische Föderation		
DE	Deutschland	LI	Liechtenstein	SD	Sudan		
DK	Dänemark	LK	Sri Lanka	SE	Schweden		
EE	Estland	LR	Liberia	SG	Singapur		

Beschreibung**Anordnung und Verfahren zur Erkennung eines vorgegebenen Wortschatzes in gesprochener Sprache durch einen Rechner**

5

Die Erfindung betrifft eine Anordnung und ein Verfahren zur Erkennung eines vorgegebenen Wortschatzes in gesprochener Sprache durch einen Rechner.

- 10 Ein Verfahren und eine Anordnung zur Erkennung gesprochener Sprache sind aus [1] bekannt. Bei der Erkennung gesprochener Sprache werden, insbesondere bis zum Erhalt einer erkannten Wortfolge aus einem digitalisierten Sprachsignal, eine Signalanalyse und eine globale Suche, die auf ein akustisches
- 15 Modell und ein linguistisches Modell der zu erkennenden Sprache zurückgreift, durchgeführt. Ein akustisches Modell basiert auf einem Phoneminventar, das anhand von Hidden-Markov-Modellen (HMMs) realisiert ist. Während der globalen Suche werden für Merkmalsvektoren, die aus der Signalanalyse
- 20 hervorgegangen sind, mit Hilfe des akustischen Modells eine passende Wortfolge ermittelt und diese als erkannte Wortfolge ausgegeben. Die zu erkennenden Wörter sind in einem Aussprachelexikon zusammen mit einer phonetischen Umschrift abgespeichert. Der Zusammenhang ist ausführlich in [1]
- 25 dargestellt.

Zur Erläuterung der nachfolgenden Ausführungen wird an dieser Stelle kurz auf die verwendeten Begriffe eingegangen.

- 30 Die Signalanalyse als Phase der computerbasierten Spracherkennung umfaßt insbesondere eine Fouriertransformation des digitalisierten Sprachsignals und eine sich daran anschließende Merkmalsextraktion. Aus [1] geht hervor, daß die Signalanalyse alle zehn Millisekunden
- 35 erfolgt. Aus sich überlappenden Zeitabschnitten mit einer Dauer von z.B. jeweils 25 Millisekunden werden anhand der Signalanalyse ungefähr 30 Merkmale ermittelt und zu einem

Merkmalsvektor zusammengefaßt. Die Komponenten des Merkmalsvektors beschreiben die spektrale Energieverteilung des zugehörigen Signalausschnitts. Um diese Energieverteilung zu erhalten, wird auf jedem Signalabschnitt (25ms-
5 Zeitabschnitt) eine Fouriertransformation durchgeführt. Aus der Darstellung des Signals im Frequenzbereich resultieren die Komponenten des Merkmalsvektors. Nach der Signalanalyse liegt das digitalisierte Sprachsignal in Form von Merkmalsvektoren vor.

10

Diese Merkmalsvektoren werden der globalen Suche, einer weiteren Phase der Spracherkennung, zugeführt. Wie bereits erwähnt, bedient sich die globale Suche des akustischen Modells und ggf. des linguistischen Modells, um die Folge von
15 Merkmalsvektoren auf Einzelteile der als Modell vorliegenden Sprache (Vokabular) abzubilden. Eine Sprache setzt sich aus einer vorgegebenen Anzahl vom Lauten, sog. Phonemen, zusammen, deren Gesamtheit als Phoneminventar bezeichnet wird. Das Vokabular wird durch Phonemfolgen modelliert und in
20 einem Aussprachelexikon abgespeichert. Jedes Phonem wird durch mindestens ein HMM modelliert. Mehrere HMMs ergeben einen stochastischen Automaten, der Zustände und Zustandsübergänge (Transitionen) umfaßt. Mit HMMs läßt sich der zeitliche Ablauf des Auftretens bestimmter
25 Merkmalsvektoren (selbst innerhalb eines Phonems) modellieren. Ein entsprechendes Phonem-Modell umfaßt dabei eine vorgegebene Anzahl von Zuständen, die linear hintereinander angeordnet sind. Ein Zustand eines HMMs stellt einen Teil eines Phonems (bspw. mit einer Dauer von 10ms)
30 dar. Jeder Zustand ist verknüpft mit einer Emissionswahrscheinlichkeit, die insbesondere nach Gauß verteilt ist, für die Merkmalsvektoren und mit Transitionswahrscheinlichkeiten für die möglichen Übergänge. Mit der Emissionsverteilung wird einem Merkmalsvektor eine
35 Wahrscheinlichkeit zugeordnet, mit der dieser Merkmalsvektor in einem zugehörigen Zustand beobachtet wird. Die möglichen Übergänge sind ein direkter Übergang von einem Zustand in

einen nächsten Zustand, ein Wiederholen des Zustands und ein Überspringen des Zustands.

Eine Aneinanderreihung von HMM-Zustände mit den zugehörigen
5 Übergängen über die Zeit wird als Trellis bezeichnet. Um die
akustische Wahrscheinlichkeit eines Wortes zu bestimmen,
verwendet man insbesondere das Prinzip der dynamischen
Programmierung: Es wird der Pfad durch die Trellis gesucht,
der den geringsten Fehler aufweist bzw. der durch die größte
10 Wahrscheinlichkeit für ein zu erkennendes Wort bestimmt ist.

Das Ergebnis der globalen Suche ist die Ausgabe bzw.
Bereitstellung einer erkannten Wortfolge, die sich unter
Berücksichtigung des akustischen Modells (Phoneminventar) für
15 jedes einzelne Wort und des Sprachmodells für die Abfolge von
Wörtern ergibt.

Aus [2] ist ein Verfahren zur Sprecheradaption, basierend auf
einer MAP-Schätzung (MAP = maximum a posteriori) von HMM-
20 Parametern bekannt.

So ist es laut [2] anerkannt, daß ein sprecherabhängiges
System zur Spracherkennung normalerweise bessere Ergebnisse
als ein sprecherunabhängiges System liefert, sofern
ausreichend Trainingsdaten verfügbar sind, die eine
25 Modellierung des sprecherabhängigen Systems ermöglichen.
Sobald jedoch die Menge der sprecherspezifischen
Trainingsdaten beschränkt ist, erreicht das
sprecherunabhängige System die besseren Resultate. Eine
Möglichkeit zur Leistungssteigerung beider Systeme, also
30 sowohl des sprecherabhängigen als auch des
sprecherunabhängigen Systems zur Spracherkennung, besteht
darin, die vorab gespeicherten Datensätze mehrerer Sprecher,
derart zu benutzen, daß auch eine kleine Menge Trainingsdaten
ausreicht, um einen neuen Sprecher in ausreichender Qualität
35 zu modellieren. Solch ein Trainingsverfahren wird
Sprecheradaption genannt. In [2] wird insbesondere die

Sprecheradaption durch eine MAP-Schätzung der Hidden-Markov-Modell-Parameter durchgeführt.

5 Generell verschlechtern sich Ergebnisse eines Verfahrens zur
Erkennung gesprochener Sprache, sobald charakteristische
Merkmale der gesprochenen Sprache von charakteristischen
10 Merkmalen der Trainingsdaten abweichen. Beispiele für
charakteristische Merkmale sind Sprechereigenschaften oder
akustische Kontexte, die sich in Form von Verschleifungen auf
die Artikulation der Phoneme auswirken.

Der in [2] verfolgte Ansatz zur Sprecheradaption besteht
darin, Parameterwerte der Hidden-Markov-Modelle
"nachzuschätzen", wobei diese nach Verarbeitung "offline",
15 d.h. nicht zur Laufzeit des Verfahrens zur Spracherkennung,
durchgeführt wird.

Die Aufgabe der Erfindung besteht darin, eine Anordnung und
ein Verfahren zur Erkennung eines vorgegebenen Wortschatzes
20 in gesprochener Sprache anzugeben, wobei insbesondere eine
Anpassung des akustischen Modells zur Laufzeit (also
"Online") vollzogen wird.

25 Diese Aufgabe wird gemäß den Merkmalen der unabhängigen
Patentansprüche gelöst.

Zur Lösung der Aufgabe wird ein Verfahren zur Erkennung eines
vorgegebenen Wortschatzes in gesprochener Sprache durch einen
Rechner angegeben, in dem aus der gesprochenen Sprache ein
30 Sprachsignal bestimmt wird. Das Sprachsignal wird einer
Signalanalyse unterworfen, woraus Merkmalsvektoren zur
Beschreibung des digitalisierten Sprachsignals hervorgehen.
Eine globale Suche wird zur Abbildung der Merkmalsvektoren
auf eine in modellierter Form vorliegende Sprache
35 durchgeführt, wobei jedes Phonem der Sprache durch ein
modifiziertes Hidden-Markov-Modell und jeder Zustand des
modifizierten Hidden-Markov-Modells durch eine

Wahrscheinlichkeitsdichtefunktion beschrieben wird. Es erfolgt eine Anpassung der Wahrscheinlichkeitsdichtefunktion derart, daß sie in eine erste Wahrscheinlichkeitsdichtefunktion und in eine zweite Wahrscheinlichkeitsdichtefunktion aufgespalten wird. Schließlich wird von der globalen Suche eine erkannte Wortfolge bereitgestellt.

Hierbei sei angemerkt, daß die Wahrscheinlichkeitsdichtefunktion, die in eine erste und in eine zweite Wahrscheinlichkeitsdichtefunktion aufgespalten wird, eine Emissionsverteilung für einen vorgegebenen Zustand des modifizierten Hidden-Markov-Modells darstellen kann, wobei diese Emissionsverteilung auch eine Überlagerung mehrerer Wahrscheinlichkeitsdichtefunktionen, z.B. Gauß-Kurven (Gauß'sche Wahrscheinlichkeitsdichteverteilungen), enthalten kann.

Eine erkannte Wortfolge kann dabei auch einzelne Lauten bzw. nur ein einzelnes Wort umfassen.

Sollte im Rahmen der globalen Suche eine Erkennung mit einem hohen Wert für den Abstand zwischen gesprochener Sprache und von der globalen Suche ermittelten dazugehöriger Wortfolge behaftet sein, so kann die Zuordnung eines Nullwortes erfolgen, welches Nullwort anzeigt, das die gesprochene Sprache nicht mit ausreichender Güte erkannt wird.

Es ist ein Vorteil der Erfindung, durch die Aufspaltung der Wahrscheinlichkeitsdichtefunktion in einem durch die Merkmalsvektoren aufgespannten Merkmalsraum neue Bereiche zu schaffen, die signifikante Information in Bezug auf die zu erkennenden digitalisierten Sprachdaten aufweisen und damit eine verbesserte Erkennung zu gewährleisten.

Eine Ausgestaltung besteht darin, daß die Wahrscheinlichkeitsdichtefunktion in die erste und in die zweite Wahrscheinlichkeitsdichtefunktion aufgespalten wird,

falls der Abfall eines Entropiewertes unterhalb einer vorgegebenen Schranke liegt.

Die Aufspaltung der Wahrscheinlichkeitsdichtefunktion in Abhängigkeit von einem Entropiewert erweist sich in der Praxis als äußerst vorteilhaft.

Die Entropie ist allgemein ein Maß für eine Unsicherheit bei einer Vorhersage eines statistischen Ereignisses. Die Entropie ist insbesondere mathematisch bestimmbar für Gauß-Verteilungen, wobei eine direkte logarithmische Abhängigkeit zwischen der Streuung σ und der Entropie besteht.

Eine andere Ausgestaltung der Erfindung besteht darin, daß die Wahrscheinlichkeitsdichtefunktionen, insbesondere die erste und die zweite Wahrscheinlichkeitsdichtefunktion jeweils mindestens eine Gauß-Verteilung umfassen.

Die Wahrscheinlichkeitsdichtefunktion des Zustandes wird durch eine Summe mehrerer Gaußverteilungen angenähert. Die einzelnen Gaußverteilungen werden Moden genannt. Bei dem vorgestellten Verfahren werden die Moden insbesondere isoliert voneinander betrachtet. Bei jedem einzelnen Aufspaltvorgang wird eine Mode in zwei Moden aufgeteilt. Wenn die Wahrscheinlichkeitsdichtefunktion aus M Moden gebildet wurde, so wird sie nach dem Aufspaltvorgang aus $M+1$ Moden gebildet. Wird eine Mode beispielsweise als eine Gaußverteilung angenommen, so kann eine Entropie berechnet werden, wie im Ausführungsbeispiel gezeigt wird.

Eine Online-Adaption ist deshalb vorteilhaft, weil das Verfahren nach wie vor Sprache erkennt, ohne in einer gesonderten Trainingsphase auf die Veränderung des Wortschatzes eingestellt werden zu müssen. Es erfolgt eine Selbstadaption, die insbesondere notwendig wird durch eine veränderte Koartikulation der Sprecher aufgrund eines Hinzufügens eines neuen Wortes.

Die Online-Adaption erfordert demnach keine gesonderte Berechnung der Wahrscheinlichkeitsdichtefunktionen, die wiederum für eine Nicht-Verfügbarkeit des Systems zur
5 Spracherkennung verantwortlich wäre.

Eine Weiterbildung der Erfindung besteht darin, daß für die erste Wahrscheinlichkeitsdichtefunktion und für die zweite Wahrscheinlichkeitsdichtefunktion gleiche
10 Standardabweichungen bestimmt werden. Ein erster Mittelwert der ersten Wahrscheinlichkeitsdichtefunktion und ein zweiter Mittelwert der zweiten Wahrscheinlichkeitsdichtefunktion werden derart bestimmt, daß der erste Mittelwert von dem zweiten Mittelwert verschieden ist.

15 Dies ist ein Beispiel für die Gewichtung der aus der Wahrscheinlichkeitsdichtefunktion aufgespaltenen ersten und zweiten Wahrscheinlichkeitsdichtefunktion. Es sind auch beliebig andere Gewichtungen vorstellbar, die auf den
20 jeweiligen Anwendungsfall anzupassen sind.

Schließlich ist es eine Weiterbildung, daß das Verfahren mehrfach hintereinander durchgeführt wird und somit eine wiederholte Aufspaltung der Wahrscheinlichkeitsdichtefunktion
25 erfolgt.

Weiterbildungen der Erfindung ergeben sich aus den abhängigen Ansprüchen.

30 Eine andere Lösung der Aufgabe besteht darin, eine Anordnung mit einer Prozessoreinheit anzugeben, welche Prozessoreinheit derart eingerichtet ist, daß folgende Schritte durchführbar sind:

35 a) aus der gesprochenen Sprache wird ein digitalisiertes Sprachsignal bestimmt; ,

- b) auf dem digitalisierten Sprachsignal erfolgt eine Signalanalyse, woraus Merkmalsvektoren zur Beschreibung des digitalisierten Sprachsignals hervorgehen;
- 5 c) eine globale Suche zur Abbildung der Merkmalsvektoren erfolgt auf eine in modellierter Form vorliegende Sprache, wobei Phoneme der Sprache durch ein modifiziertes Hidden-Markov-Modell und jeder Zustand des Hidden-Markov-Modells durch eine
- 10 Wahrscheinlichkeitsdichtefunktion beschreibbar ist;
- d) die Wahrscheinlichkeitsdichtefunktion durch Veränderung des Wortschatzes angepaßt, indem die Wahrscheinlichkeitsdichtefunktion in eine erste Wahrscheinlichkeitsdichtefunktion und in eine zweite
- 15 Wahrscheinlichkeitsdichtefunktion aufgespalten wird;
- e) von der globalen Suche wird eine erkannte Wortfolge bereitgestellt.

Diese Anordnung ist insbesondere geeignet zur Durchführung des erfindungsgemäßen Verfahrens oder einer seiner vorstehend erläuterten Weiterbildungen.

20

Ausführungsbeispiele der Erfindung werden nachfolgend anhand der Zeichnung dargestellt und erläutert.

25

Es zeigt

Fig.1 eine Anordnung bzw. ein Verfahren zur Erkennung gesprochener Sprache.

30

In Figur 1 sind eine Anordnung bzw. ein Verfahren zur Erkennung gesprochener Sprache dargestellt. Zur Erläuterung der nachstehend verwendeten Begriffe sei auf die Beschreibungseinleitung verwiesen.

35

Ein digitalisiertes Sprachsignal 101 wird in einer Signalanalyse 102 einer Fouriertransformation 103 mit

anschließender Merkmalsextraktion 104 unterzogen. Die Merkmalsvektoren 105 werden an ein System zur globalen Suche 106 übermittelt. Die globale Suche 106 berücksichtigt sowohl ein akustisches Modell 107 als auch ein linguistisches Modell 108 zur Bestimmung der erkannten Wortfolge 109. Aus dem digitalisierten Sprachsignal 101 geht somit die erkannte Wortfolge 109 hervor.

In dem akustischen Modell 107 wird das Phoneminventar anhand von Hidden-Markov-Modellen nachgebildet.

Eine Wahrscheinlichkeitsdichtefunktion eines Zustands des Hidden-Markov-Modells wird durch eine Aufsummierung einzelner Gaußscher Moden angenähert. Eine Mode ist insbesondere eine Gaußglocke. Durch Aufsummierung mehrerer Moden entsteht eine Mischung einzelner Gaußglocken und damit eine Modellierung der Emissionswahrscheinlichkeitsdichtefunktion. Anhand eines statistischen Kriteriums wird entschieden, ob der zu erkennende Wortschatz des Spracherkenners durch das Hinzufügen weiterer Moden verbessert modelliert werden kann. Im Fall der vorliegenden Erfindung wird dies insbesondere bei Erfüllung des statistischen Kriteriums durch inkrementelles Aufspalten bereits existierender Moden erreicht.

Die Entropie ist definiert durch

$$H_p = - \int_{-\infty}^{\infty} p(\bar{x}) \log_2 p(\bar{x}) d\bar{x} \quad (1)$$

unter der Annahme, daß $p(\bar{x})$ eine Gauß-Verteilung mit einer diagonalen Kovarianzmatrix ist, also

$$p(\bar{x}) = \mathcal{N}(\bar{\mu}, \sigma_n) = \frac{1}{\sqrt{(2\pi)^N}} \frac{1}{\prod_n \sigma_n} \cdot \exp\left(-\frac{1}{2} \sum_n \frac{(x_n - \mu_n)^2}{\sigma_n^2}\right) \quad (2)$$

erhält man

$$H_p = \sum_{n=1}^N \log_2 \sqrt{2\pi e} \sigma_n \quad (3),$$

5 wobei

μ den Erwartungswert,
 σ_n die Streuung für jede Komponente n und
 N die Dimension des Merkmalsraums

bezeichnen.

10

Die wahre Verteilung $p(\bar{x})$ ist nicht bekannt. Sie wird insbesondere als Gaußverteilung angenommen. Im akustischen Modell wird die Wahrscheinlichkeit $p(\bar{x})$ anhand von Stichproben angenähert mit

15

$$\hat{p}(\bar{x}) = \mathcal{N}(\bar{\mu}, \sigma_n),$$

wobei

20

$$\bar{\mu} = \frac{1}{L} \sum_{l=1}^L \bar{x}_l$$

einen Mittelwert über L Beobachtungen darstellt. Die korrespondierende Entropie als Funktion von $\hat{\mu}$ ist gegeben durch

25

$$H_{\hat{p}}(\hat{\mu}) = - \int_{-\infty}^{\infty} p(\bar{x}) \log_2 \hat{p}(\bar{x}) d\bar{x} \quad (4),$$

was schließlich zu

30

$$H_{\hat{p}}(\hat{\mu}) = H_p + \sum_{n=1}^N \frac{(\mu_n - \hat{\mu}_n)^2}{\sigma_n^2} \log_2 \sqrt{e} \quad (5)$$

führt.

Der Erwartungswert $E\{(\mu_n - \hat{\mu}_n)^2\}$ beträgt $\frac{1}{L} \sigma_n^2$, so daß der Erwartungswert von $H_{\hat{p}}(\hat{\mu})$ gegeben ist als

5

$$H_{\hat{p}} = E\{H_{\hat{p}}(\hat{\mu})\} = H_p + \frac{N}{L} \log_2 \sqrt{e} \quad (6).$$

Für die Entropie einer Mode, die mit einer Gauß-Verteilung mit einer diagonalen Kovarianzmatrix bestimmt wird, ergibt sich also Gleichung (3). Der Prozeß wird nun mit einer Schätzung angenähert. Die Entropie des angenäherten Prozesses ergibt sich zu

$$\hat{H} = H + \frac{N}{L} \log_2 \sqrt{e} \quad (7).$$

15

Je größer die Anzahl L der Stichproben ist, um so besser wird die Abschätzung und um so mehr nähert sich die geschätzte Entropie \hat{H} der wahren Entropie H an.

20 Es soll nun

$$p(\bar{x}) = \mathcal{N}(\bar{\mu}, \sigma_n) \quad (8)$$

die aufzuteilende Mode sein. Ferner wird angenommen, daß die zwei Gauß-Verteilungen, die durch den Aufteilungsprozeß entstehen, identische Standardabweichungen σ^S haben und gleich gewichtet sind. Dies ergibt

25

$$p^S(\bar{x}) = \frac{1}{2} \mathcal{N}(\bar{\mu}_1^S, \sigma^S) + \frac{1}{2} \mathcal{N}(\bar{\mu}_2^S, \sigma^S) \quad (9).$$

30

Unter der Annahme, daß $\mu_1 \approx \hat{\mu}_1$, $\mu_2 \approx \hat{\mu}_2$ und daß μ_1 ausreichend weit entfernt von μ_2 ist, ergibt sich die Entropie der aufgespaltenen Wahrscheinlichkeitsdichtefunktion jeweils zu

$$\hat{H}^S = 1 - \sum_{n=1}^N \log_2 \sqrt{2\pi e} \sigma_n^S + \frac{1}{2} \left(\log_2 \sqrt{e} \frac{N}{L_1} + \log_2 \sqrt{e} \frac{N}{L_2} \right) \quad (10).$$

Als Aufteilungskriterium wird eine Verminderung der Entropie
5 durch den Aufspaltungsvorgang gefordert, also

$$\hat{H} - \hat{H}^S > C \quad (11),$$

wobei C (mit $C > 0$) eine Konstante ist, die den gewünschten
10 Abfall der Entropie darstellt. Wird

$$\frac{L}{2} = L_1 = L_2 \quad (12)$$

angenommen, so ergibt sich hierdurch

15

$$\sum_{n=1}^N \log_2 \frac{\sigma_n}{\sigma_n^S} > \log_2 \sqrt{e} \frac{N}{L} + 1 + C \quad (13).$$

Eine Möglichkeit, die Lage der Mittelpunkte der beiden neuen
Moden zu bestimmen, wird im folgenden aufgezeigt. Eine
20 bevorzugte Vorgabe besteht darin, daß Kriterium zum
Aufspalten zu erfüllen. In dem angeführten Beispiel wird $\hat{\mu}_1^S$
der Wert von $\hat{\mu}$ zugewiesen. $\hat{\mu}_2^S$ erhält eine Maximum-
Likelihood-Schätzung derjenigen Beobachtungen, die im
Viterbi-Pfad auf $\hat{\mu}$ abgebildet werden. Diese Bestimmungen
25 zeigen lediglich eine Möglichkeit auf, ohne daß eine
Einschränkung des vorgestellten Verfahrens auf diese
Möglichkeit beabsichtigt ist.

Die folgenden Schritte der Beispielanwendung zeigen die
30 Einbettung in eine Anordnung zur Spracherkennung bzw. ein
Verfahren zur Spracherkennung.

Schritt 1: Initialisierung: $\bar{\mu}_1^S = \bar{\mu}$, $\bar{\mu}_2^S = \bar{\mu}$.

Schritt 2: Erkennen der Äußerung , Analysieren des Viterbi-Pfads;

5 Schritt 3: Für jeden Zustand und für jede Mode des Viterbi-Pfades:

Schritt 3.1: Bestimme σ_n ;

10 Schritt 3.2: Bestimme L_2 auf Grundlage derjenigen Beobachtungen, die näher an $\bar{\mu}_2^S$ als an $\bar{\mu}_1^S$ liegen und setze $L = L_2$. Falls $\bar{\mu}_2^S$ und $\bar{\mu}_1^S$ identisch sind, so ordne die zweite Hälfte der Merkmalsvektoren $\bar{\mu}_2^S$ und die
15 erste Hälfte der Merkmalsvektoren $\bar{\mu}_1^S$ zu.

Schritt 3.3: Bestimme σ_n^S entsprechend auf Grundlage der L_2 -Äußerungen;

20 Schritt 3.4: Ermittle $\bar{\mu}_2^S$ neu auf Grundlage des Mittelwerts derjenigen Beobachtungen, die näher an $\bar{\mu}_2^S$ als an $\bar{\mu}_1^S$ liegen;

25 Schritt 3.5: Werte Aufteilungskriterium nach Gleichung (13) aus;

Schritt 3.6: Falls Aufteilungskriterium nach Gleichung (13) positiv ist, generiere zwei neue Moden mit den Mittelpunkten $\bar{\mu}_1^S$
30 und $\bar{\mu}_2^S$.

Schritt 4: Gehe zu Schritt 2.

Im Rahmen dieses Dokuments wurden folgende Veröffentlichungen zitiert:

- 5 [1] N. Haberland et al.: "Sprachunterricht - Wie funktioniert die computerbasierte Spracherkennung?", c't - Magazin für Computertechnik - 5/1998, Heinz Heise Verlag, Hannover, 1998, Seiten 120 bis 125.
- 10 [2] C. H. Lee et al.: "Speaker Adaptation Based on MAP Estimation of HMM Parameters"; Proc. IEEE Intern. Conference on Acoustics, Speech and Signal Processing, Seiten II-588 bis II-561.

Patentansprüche

1. Verfahren zur Erkennung eines vorgegebenen Wortschatzes
in gesprochener Sprache durch einen Rechner,
 - a) bei dem aus der gesprochenen Sprache ein
digitalisiertes Sprachsignal bestimmt wird,
 - b) bei dem auf dem digitalisierten Sprachsignal eine
Signalanalyse durchgeführt wird, woraus
Merkmalsvektoren zur Beschreibung des digitalisierten
Sprachsignals hervorgehen,
 - c) bei dem eine globale Suche zur Abbildung der
Merkmalsvektoren auf eine in modellierter Form
vorliegende Sprache durchgeführt wird, wobei Phoneme
der Sprache durch ein modifiziertes Hidden-Markov-
Modell und jeder Zustand des Hidden-Markov-Modells
durch eine Wahrscheinlichkeitsdichtefunktion
beschrieben wird,
 - d) bei dem die Wahrscheinlichkeitsdichtefunktion durch
Veränderung des Wortschatzes angepaßt wird, indem die
Wahrscheinlichkeitsdichtefunktion in eine erste
Wahrscheinlichkeitsdichtefunktion und in eine zweite
Wahrscheinlichkeitsdichtefunktion aufgespalten wird,
 - e) bei dem von der globalen Suche eine erkannte Wortfolge
bereitgestellt wird.
2. Verfahren nach Anspruch 1,
bei dem die Wahrscheinlichkeitsdichtefunktion in die
erste und in die zweite Wahrscheinlichkeitsdichtefunktion
aufgespalten wird, falls der Abfall eines Entropiewertes
unterhalb einer vorgegebenen Schranke liegt.
3. Verfahren nach Anspruch 1 oder 2,
bei dem die Veränderung des Wortschatzes zur Laufzeit des
Verfahrens durchgeführt wird.
4. Verfahren nach einem der vorhergehenden Ansprüche,
bei dem die Veränderung des Wortschatzes bedingt ist

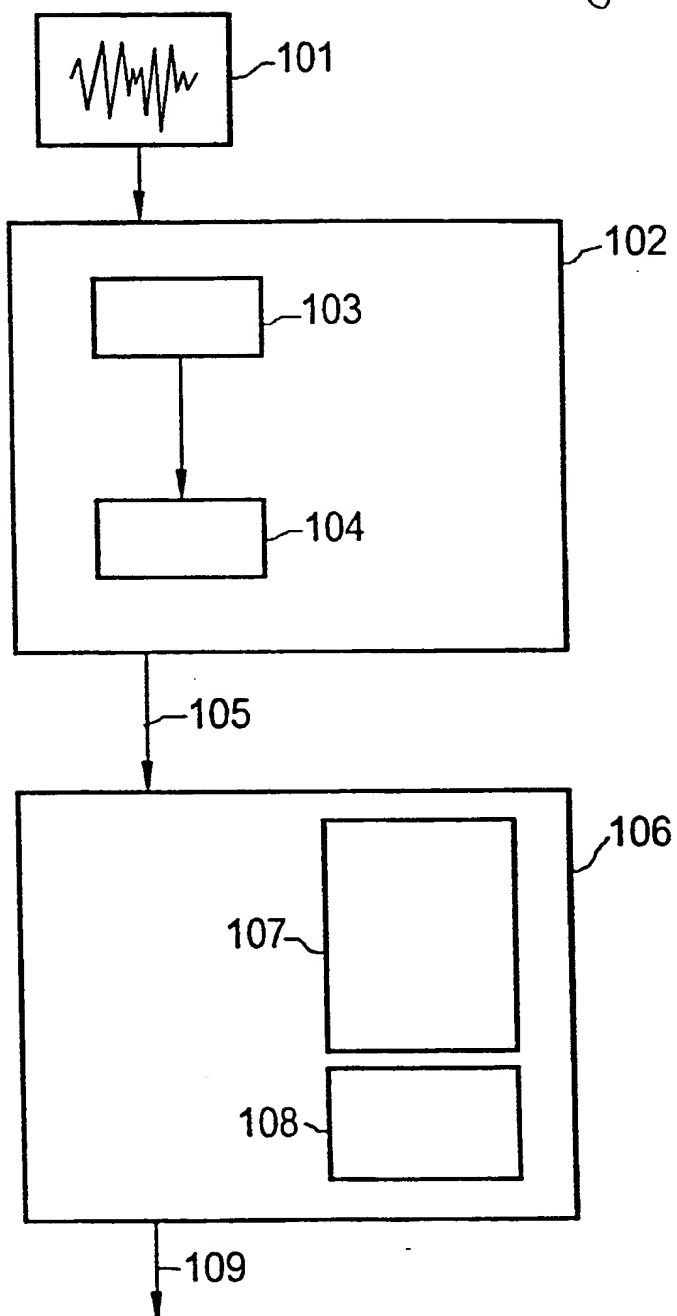
durch Hinzufügen eines Wortes zum Wortschatz oder bei dem sich Aussprachegewohnheiten eines Sprechers ändern.

5. Verfahren nach einem der vorhergehenden Ansprüche,
5 bei dem die erste Wahrscheinlichkeitsdichtefunktion und die zweite Wahrscheinlichkeitsdichtefunktion jeweils mindestens eine Gaußverteilung umfassen.
6. Verfahren nach Anspruch 5,
10 bei dem für die erste Wahrscheinlichkeitsdichtefunktion und für die zweite Wahrscheinlichkeitsdichtefunktion gleiche Standardabweichungen, ein erster Mittelwert der ersten Wahrscheinlichkeitsdichtefunktion und ein zweiter Mittelwert der zweiten Wahrscheinlichkeitsdichtefunktion
15 ermittelt werden, wobei der erste Mittelwert von dem zweiten Mittelwert verschieden ist.
7. Verfahren nach einem der vorhergehenden Ansprüche,
20 bei dem die Aufspaltung mehrfach durchgeführt wird.
8. Anordnung zur Erkennung eines vorgegebenen Wortschatzes in gesprochener Sprache mit einer Prozessoreinheit, die derart eingerichtet ist, daß
25 a) aus der gesprochenen Sprache ein digitalisiertes Sprachsignal bestimmbar ist,
b) auf dem digitalisierten Sprachsignal eine
30 Signalanalyse durchführbar ist, woraus Merkmalsvektoren zur Beschreibung des digitalisierten Sprachsignals hervorgehen,
c) eine globale Suche zur Abbildung der Merkmalsvektoren auf eine in modellierter Form vorliegende Sprache erfolgt, wobei Phoneme der Sprache durch ein
35 modifiziertes Hidden-Markov-Modell und jeder Zustand des Hidden-Markov-Modells durch eine Wahrscheinlichkeitsdichtefunktion beschreibbar ist,

- 5 d) die Wahrscheinlichkeitsdichtefunktion durch
Veränderung des Wortschatzes angepaßt wird, indem die
Wahrscheinlichkeitsdichtefunktion in eine erste
Wahrscheinlichkeitsdichtefunktion und in eine zweite
Wahrscheinlichkeitsdichtefunktion aufgespalten wird,
e) von der globalen Suche eine erkannte Wortfolge
bereitgestellt wird.

1 / 1

Fig. 1



A. KLASIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
 IPK 6 G10L5/06 G10L3/00

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)
 IPK 6 G10L

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie ^o	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	JP 09 152886 A (ATR INTERPRETING TELECOM) 10. Juni 1997 (1997-06-10) Zusammenfassung	1,8
P, A	-& US 5 839 105 A (OSTENDORF ET AL.) 17. November 1998 (1998-11-17) Spalte 4 -Spalte 6, Zeile 31 ---	1,8
X	TAKAMI ET AL.: "A successive state splitting algorithm for efficient allophone modeling" INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP) 1992, Bd. 1, 23. - 26. März 1992, Seiten 573-576, XP000341211 SAN FRANCISCO, CA, US ISBN: 0-7803-0532-9 Seite 574 -Seite 575, linke Spalte --- -/--	1,8



Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen



Siehe Anhang Patentfamilie

^o Besondere Kategorien von angegebenen Veröffentlichungen :

"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

12. Oktober 1999

Absenddatum des internationalen Recherchenberichts

19/10/1999

Name und Postanschrift der Internationalen Recherchenbehörde

Europäisches Patentamt, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
 Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Lange, J

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	<p>LEE ET AL.: "Speaker adaptation based on map estimation of HMM parameters"</p> <p>INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP) 1993,</p> <p>Bd. 2, 27. - 30. April 1993, Seiten 558-561, XP000427850</p> <p>MINNEAPOLIS, MN, US ISBN: 0-7803-0946-4</p> <p>in der Anmeldung erwähnt</p> <p>das ganze Dokument</p> <p>-----</p>	1,8

Angaben zu Veröffentlichung, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

CT/DE 99/01323

Formblatt PCT/ISA/210 (Anhang Patentfamilie)(Juli 1992)